

DOCUMENT RESUME

ED 160 599

95

SP 013 168

AUTHOR Suppes, Patrick, Ed.
TITLE Impact of Research on Education: Some Case Studies.
INSTITUTION National Academy of Education, Washington, D.C.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 78
CONTRACT NIE-C-74-0116
NOTE 679p.; For related document, see SP 013 167 ; Parts marginally legible due to small type
AVAILABLE FROM National Academy of Education, 11 Dupont Circle, Washington, D.C. 20036 (\$15.00 + \$1.50 postage)
EDRS PRICE MF-\$1.33 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Behavior Theories; *Case Studies; *Educational Improvement; Educational Legislation; Educational Policy; Educational Psychology; *Educational Research; English (Second Language); Government Role; Linguistics; Predictive Ability (Testing); *Research Utilization; School Architecture; Scientific Methodology; *Technology Transfer; Vocabulary Development

ABSTRACT

Significant cases of research in terms of their effect on school practices are cited and discussed in this volume. In presenting the various case studies, the authors have not focused on the scientific validity of the ideas they present but on the impact of these ideas on education practice. The following topics are covered: (1) the use of psychological tests, especially tests of intellectual capacity, and the controversies that surround current testing practices; (2) the impact of vocabulary research on the actual content of written materials used in schools; (3) recent approaches to behavior modification in educational settings; (4) the implications for education of the theories of Jean Piaget; (5) the influence of psychological theories in American elementary education; (6) the impact of the language sciences on second language education, and the developments in psycholinguistics and social linguistics over the past thirty years; (7) the use of theory-oriented work in the architectural design of classrooms; (8) the relation between educational research and educational reform in Sweden; and (9) the relation between social science research and the law, and the issue of school-finance reform. (JD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED160599

IMPACT OF RESEARCH ON EDUCATION: SOME CASE STUDIES

Edited by

PATRICK SUPPES

"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY

NAE

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

NATIONAL ACADEMY OF EDUCATION
Washington, D.C.

1978

\$15.00

1.50

\$16.50

2

SP013 168

The project reported herein was performed pursuant to a contract (NIE-C-74-0116) from the National Institute of Education, U.S. Department of Health, Education, and Welfare. However, the opinions expressed do not necessarily reflect NIE position or policy, and no official endorsement by the National Institute of Education or the U.S. Department of Health, Education, and Welfare should be inferred.

Business correspondence should be addressed to
National Academy of Education
11 Dupont Circle
Washington, D.C. 20036

Copyright © 1978 by National Academy of Education
All rights reserved.

Printed in the United States of America

CONTENTS

PREFACE

xiii

John B. Carroll

Chapter 1. ON THE THEORY-PRACTICE INTERFACE IN THE MEASUREMENT OF INTELLECTUAL ABILITIES 1

1. The Period of Development 4

- 1.1 Overview / 4
- 1.2 Notions of mental ability / 9
- 1.3 Theories of measurement and scaling / 14
- 1.4 Statistical methods and mental test theory / 18
- 1.5 Test construction, standardization, and validation methods / 24
- 1.6 Practices in the use of tests / 33

2. The Modern Period 37

- 2.1 Overview / 37
- 2.2 Psychometrics and theories of intellectual abilities / 44
- 2.3 Test theory in the modern period / 46
- 2.4 Factor analysis in the modern period / 50
- 2.5 Factor-analytic studies in cognitive abilities / 56
- 2.6 Theories of measurement and scaling / 59
- 2.7 Statistical and computational technology / 61
- 2.8 Practices in test construction, standardization, and validation / 62
- 2.9 The use and misuse of intelligence tests in schools / 71

3. Review and Assessment: The Current Scene 78

- 3.1 Public controversies about mental testing / 79
- 3.2 New directions for mental test research / 85
- 3.3 Theories of mental abilities / 85
- 3.4 The heredity vs. environment issue / 87
- 3.5 The problem of relevance and validity / 89
- 3.6 Other problems / 90
- 3.7 The theory-practice interface in perspective: A concluding statement / 91

References 94

Geraldine Joncich Clifford

Chapter 2. WORDS FOR SCHOOLS: THE APPLICATIONS IN EDUCATION OF THE VOCABULARY RESEARCHES OF EDWARD L. THORNDIKE 107

1. The Origins of Contemporary Interest in Objective Vocabulary Data 109

- 1.1 The issues of adult vocabulary size / 109
- 1.2 The vocabulary development of the American child / 111

1.3 Vocabulary and the "new type" testing movement / 114	
1.4 Intelligence tests and pressure for democratic education / 115	
2. Studies in Word Usages: The Word Counters	122
2.1 Studies in word frequency before 1921 / 122	
2.2 <i>The Teacher's Word Book</i> / 125	
2.3 The question of children's vocabularies / 127	
2.4 Technical vocabularies and the textbook / 129	
3. Conclusions Drawn from Frequency Research and Their Implications for Education	134
3.1 Overlapping vocabularies and curriculum integration / 134	
3.2 Overlapping adult and child vocabularies and ESL instruction / 137	
3.3 The skewed distribution curve of word frequencies and practical instruction / 140	
3.4 The characteristics of frequent words / 146	
4. On How and Why Things Changed in Schools	152
4.1 The media of Thorndike's influence / 152	
4.2 What makes research influence practice? / 165	
References	183

Robert Glaser

<i>Chapter 3. THE CONTRIBUTIONS OF B. F. SKINNER TO EDUCATION AND SOME COUNTERINFLUENCES</i>	199
1. The Relationship of Psychology to Education	199
1.1 Contrasting approaches / 201	
1.2 A loosening of ties / 202	
2. Background	204
2.1 Skinner's system of behavior / 204	
2.2 Reinforcement schedules / 206	
2.3 Kinds of reinforcers / 207	
2.4 Shaping and differentiation of responses / 208	
2.5 Discrimination / 209	
2.6 Generalization and abstraction / 210	
2.7 Complexity and optimality / 212	
2.8 Early counterinfluences / 212	
2.9 Extrapolation to human behavior / 214	
3. Teaching Machines and Programmed Instruction	215
3.1 Application of operant principles / 215	
3.2 Application to the classroom / 217	
3.3 Some details of design and implementation / 220	
4. Rapid Development	222
4.1 An insular technological enterprise / 223	
4.2 Keller: A system of personalized instruction (PSI) / 225	
4.3 Individualizing instruction / 228	
5. Behavior Modification	229
5.1 The general approach / 230	
5.2 Behavior modification in "natural" settings and token economies / 233	

CONTENTS

vii

5.3 Behavior modification in educational settings and token reinforcement in the classroom / 235	
5.4 Debate and constraints / 240	
5.5 Self-regulation / 241	
6. Counterinfluences	244
6.1 Social learning theory / 244	
6.2 Self-reinforcement: Personal standards and achievement / 247	
6.3 Covert self-reinforcement / 249	
6.4 Reinforcement and cognition / 251	
6.5 Human factors research and information processing / 252	
7. Concluding Observations	257
References	260
<i>Guy J. Groen</i>	
Chapter 4. THE THEORETICAL IDEAS OF PIAGET AND EDUCATIONAL PRACTICE	267
1. The Theory	268
1.1 The first phase / 268	
1.2 The second phase / 269	
1.3 The rediscovery of Piaget / 271	
1.4 The third phase / 278	
2. A Broader Perspective	279
2.1 Piagetian theory and scientific method / 279	
2.2 The hard core of Piagetian theory / 283	
3. Some Implications for Education	288
3.1 Piaget on education / 289	
3.2 Piagetian techniques / 292	
3.3 Stage theory / 295	
4. Some Specific Applications	299
4.1 Mathematics / 299	
4.2 Early education / 301	
4.3 The LOGO environment / 306	
5. Some Concluding Comments	311
References	312
<i>Patrick Suppes and Hermine Warren</i>	
Chapter 5. PSYCHOANALYSIS AND AMERICAN ELEMENTARY EDUCATION	319
1. Brief Historical Overview of Psychoanalysis in American Medicine and Psychology	320
1.1 Medicine / 321	
1.2 Psychology / 322	
2. Psychoanalytic Theory Pertinent to Education	324
2.1 Anna Freud's <i>Lectures to Teachers</i> / 325	
2.2 Theory of learning / 327	
2.3 Learning difficulties / 331	

2.4 Discipline and moral development / 334	
2.5 The teacher / 336	
3. Evidence of Impact	340
3.1 The child-study and child-development movements / 342	
3.2 Educational psychology / 350	
3.3 Mental hygiene / 358	
3.4 Progressive education / 371	
3.5 Cultural impact / 374	
4. A Comparison of Freud, Dewey, and Thorndike	381
5. Conclusion	385
5.1 The diffuse influence of psychoanalysis on education / 386	
5.2 The increasing influence of psychoanalysis over time / 386	
5.3 Ambivalence toward psychoanalysis / 387	
5.4 Nonprofessional channels / 387	
5.5 Difficulty of causal analysis / 388	
5.6 Limitations of education / 388	
References	388

H. H. Stern, Marjorie Bingham Wesche, and Birgit Harley

Chapter 6. THE IMPACT OF THE LANGUAGE SCIENCES ON SECOND LANGUAGE EDUCATION	397
1. Historical Orientation	401
1.1 1880 to 1920 / 402	
1.2 1920 to 1940 / 403	
1.3 1940 to 1960 / 404	
1.4 Trends: 1960 to 1975 / 407	
2. Models of Interaction	407
2.1 The separation or "hands off" model / 408	
2.2 The application model / 408	
2.3 The resource model / 409	
2.4 The common ground or convergent model / 409	
3. Linguistics and Sociolinguistics	410
3.1 The confident application of linguistics / 410	
3.2 The disorienting impact of new theory / 414	
3.3 Reassessment of the relationship / 421	
3.4 New directions / 424	
4. Psychology and Psycholinguistics	433
4.1 Changing research interests / 433	
4.2 Psychological variables related to second-language acquisition / 434	
4.3 The learner and language-learning processes / 441	
4.4 Conclusions / 447	
5. Educational Research	448
6. Mechanisms of Interaction	451
6.1 Mechanisms for initiating and carrying out research and development / 451	

CONTENTS

ix

6.2 Mechanisms of mediation and information exchange / 453	
6.3 Teacher training / 457	
7. Review and Assessment	457
7.1 The impact of linguistic theory / 458	
7.2 The impact of descriptive linguistics / 460	
7.3 The impact of psychology and psycholinguistics / 460	
7.4 The impact of educational research / 462	
7.5 Assessment of mechanisms / 462	
8. Conclusion	463
References	465

J. W. Getzels

Chapter 7. PARADIGM AND PRACTICE: ON THE IMPACT OF BASIC RESEARCH IN EDUCATION

1. Effects on Specific Classroom Practices	480
1.1 Effect on attitudes toward giftedness and programs for the gifted / 480	
1.2 Effect on leadership style and classroom environment / 482	
1.3 Effects on curriculum / 483	
1.4 Other basic research / 486	
2. Effects on Broad Educational Policy	487
2.1 Conceptions of the learner and architecture of the classroom / 488	
2.2 Empty organism: Rectangular classroom / 489	
2.3 Active organism: Square classroom / 491	
2.4 Social organism: Circular classroom / 492	
2.5 A changing paradigm / 493	
2.6 Stimulus-seeking organism: Open classroom / 495	
3. Effects on Professional Preparation in Education	497
3.1 The preparation of educational administrators / 498	
3.2 Scientific management / 498	
3.3 Human relations management / 499	
3.4 From vocational principles to theoretical concepts / 501	
3.5 Concepts and research in administration and the preparation of educational administrators / 502	
3.6 From prescriptive solutions to theoretical inquiries / 503	
4. Basic Research and the University	506
4.1 The identification of basic problems / 507	
4.2 The question of accountability / 508	
5. Theory, Research, and Practice	509
5.1 The question of theoretical bias / 510	
5.2 Consequences of avoiding theory / 511	
5.3 Objections to utilizing theory / 513	
5.4 Theoretical safeguards / 514	
6. Summary	515
References	517

Torsten Husén

Chapter 8. EDUCATIONAL RESEARCH AND EDUCATIONAL REFORM:	523
A CASE STUDY OF SWEDEN	
1. The Framework for Policymaking	523
2. Assessing Reserves of Ability and Providing Equality of Opportunity for Advanced Education	531
3. Differentiation, Individual Differences, and Disparities in Educational Opportunities	541
3.1 Structure of compulsory schooling / 541	
3.2 The problem of differentiation in Swedish educational reform, 1940-1960 / 543	
3.3 Differentiation and social background / 549	
4. Objectives and Content of Curricula	555
4.1 Studies for the 1957 School Commission / 555	
4.2 Aims and limitations of the studies / 556	
4.3 Assessing course content in Grades 7 through 10 / 558	
4.4 Assessing admission requirements of secondary schools / 559	
4.5 Assessing the requirements of employment and the individual needs for specific skills / 560	
4.6 Establishment of reference norms / 561	
4.7 Results in mathematics / 562	
4.8 The need for mathematical knowledge for further vocational training and continued academic schooling / 563	
4.9 The need for mathematical skills in various vocations and in everyday activities / 564	
4.10 Retention of skills at graduation and some years afterwards / 564	
4.11 Implications for the mathematics curriculum / 565	
4.12 The communication skills as a prerequisite for vocational training and continued academic schooling / 567	
4.13 The need for communication skills in selected occupational areas and in everyday activities / 568	
4.14 Retention of skills at graduation and some years later / 569	
4.15 Implications for curriculum revision in Swedish / 570	
4.16 Inquiries into <i>Gymnasium</i> (pre-university) education / 571	
4.17 Some observations in retrospect / 574	
References	577

Marvin Bressler and Joel S. Berke

Chapter 9. EDUCATION, SOCIAL SCIENCE, AND THE LAW:	581
THE RODRIGUEZ CASE AND SCHOOL FINANCE REFORM	
1. The <i>Rodriguez</i> Case: A Synopsis	582
2. The Study of School Finance: Advocacy and Eclecticism	583
2.1 The early tradition of scholarship in school finance / 584	
2.2 The influence of economics on school finance / 586	
2.3 Legal scholarship and the reform of school finance / 588	
3. <i>San Antonio v. Rodriguez</i> : Law and Social Science	591
3.1 The judicial resolution of educational issues / 593	

CONTENTS

xi

4. Social Science, the Law, and Educational Reform: Some Observations on Judicial Reasoning	606
4.1 The sources of authority / 606	
4.2 The criterion of relevance / 607	
4.3 The limitations of traditional social science in the judicial process / 608	
5. Attorneys and Social Scientists: Partners in Advocacy	612
6. Recent Developments in Social Advocacy, Policy Research, and Public Policy on School Finance	615
6.1 Recent legal developments / 615	
6.2 Contemporary school-finance research / 617	
6.3 Contemporary legislation to reform school finance / 619	
APPENDIX A: The Evidentiary Affidavit Submitted by Joel S. Berke in <i>Rodriguez v. San Antonio</i>	621
APPENDIX B: Excerpts from the Opinions in <i>San Antonio v. Rodriguez</i> of Justices Powell and Marshall on Data Relevant to District and Personal Poverty	640
References	643
NAME INDEX	647
SUBJECT INDEX	659

PREFACE

EVERY modern society has as a major concern the education of its citizens. In some developing countries the educational activities of the government consume as much as a third of the national budget. In the United States, such activities currently require more than a hundred billion dollars a year, although, as in the case of other forms of social and economic activity, only a slight effort is made to study the character of these educational activities and to understand them as intellectual, economic, or social processes. It is true that there has been a long tradition, even if a fragile one, of studying educational matters, but most members of the educational community will agree that educational research is a relatively minor activity in comparison to education as a whole.

All of us on occasion probably feel that there is little hope that research, given the small national effort devoted to it, will seriously affect practice. Such pessimism is not historically supported by the evidence, however, and I digress for a moment to examine some examples of how great an impact scientific research can have on society.

All of the characteristic features of modern electronic communication and rapid transportation are unique products of the long tradition of science and technology. The changes that have taken place recently, e.g., the widespread introduction of color television, have especially depended on prior scientific research. Eight outstanding recent cases that have been studied (Battelle Report, National Science Foundation, Washington, D.C., 1973) give a good sense of the diversity of important contributions to society arising from specific scientific work. The eight cases represent developments that almost certainly would not have taken place on the basis of either enlightened common sense or bare empiricism, and they range across a variety of scientific theories and domains of application: the heart pacemaker; the development of hybrid grains and the green revolution; electrophotography, which led to office copiers; input-output economic analysis developed originally in the 1930s by Leontief; organophosphorus insecticides; oral contraceptives, which rest on relatively delicate matters of steroid chemistry; magnetic ferrites, which are widely used in communications equipment and computers; and videotape recorders, which depended upon a union of electromagnetic and communication theory with the technology of audio recording. Compared with the practical significance of these scientific and technological developments, the initial cost of research and development was relatively minor.

Research can have an impact in our society and it certainly does in many different ways. In the case of education, every large school system

has as part of its central office staff some sort of research unit. The schools and colleges of education associated with institutions of higher education throughout the country are all charged with research responsibilities, and some of these are specifically written into the legislative charter of the institution. When the Office of Education was established by federal legislation more than a hundred years ago in 1867, the first section of the Act defined its chief purpose as one of "collecting such statistics and facts as shall show the condition and progress of education in the several states and territories, and of diffusing information respecting the organization and management of schools and school systems and methods of teaching."

In the hundred years since the founding of the Office of Education, a great deal of educational research has taken place, and much of it has been sponsored by the federal government. In a previous study entitled *Research for Tomorrow's Schools* (L. J. Cronbach and P. Suppes, Eds., New York, Macmillan, 1969) a committee on educational research of the National Academy of Education outlined some significant cases of research in terms of their effect on school practices. One example was the theory of mental tests and pupil classification; and the ways in which the theory has come to be applied in American public schools; another concerned the influence of Peirce's and Dewey's concepts of knowledge on the formulation of educational aims; a third was concerned with the contributions of Thorndike's theory of learning to the teaching of arithmetic in the early decades of this century.

A number of members of the National Academy of Education have expressed the view that it would be appropriate to embark on a more systematic presentation of the kinds of examples just cited. The present work is devoted to nine such studies. These studies have a variety comparable to that of the Battelle list mentioned above, but, to use a distinction that has become familiar in discussions of modern science and technology, the examples studied primarily concern "software" rather than "hardware." Certainly this is what is to be expected in the analysis of the contributions of research to educational practice.

In presenting the various case studies the authors have not focused on the scientific validity of the ideas they present, but on the impact of these ideas on educational practice. Without doubt, searching questions about validity could be raised, as is always the case in science to one degree or another, but at a time when the relevance of basic research to practice is being challenged, or at least not strongly supported in many quarters, an examination of the contributions to education of earlier research seems timely.

In Chapter 1, John B. Carroll examines in some detail the development and present status of the scientific knowledge that lies back of the extensive use of psychological tests, especially tests of intellectual capacity. Carroll's detailed and objective examination of both theory and use of

tests that measure intellectual abilities provides a desirable perspective on the many controversies that surround current testing practices.

In Chapter 2, Geraldine Joncich Clifford examines the impact of Edward L. Thorndike's vocabulary research on the actual content of written material used in schools. From a historical standpoint, Thorndike's work is a paradigm example of scientific research that has influenced detailed matters of curriculum. Clifford casts light on several factors that tend to promote or retard educational innovation as well as on the problems that arise in implementing research findings in school practice.

In Chapter 3, Robert Glaser examines the contributions of B. F. Skinner to education. That Skinner's research has had an impact on education in this country, especially in the decade and a half after World War II, is widely recognized, but the broad intellectual perspective that Glaser provides on Skinner's work and its implications, especially on more recent approaches to behavior modification in educational settings, will be found clarifying, even by those who consider themselves close to the subject.

In Chapter 4, Guy J. Groen examines the theoretical ideas of Piaget and their contributions to educational practice. Groen's lucid account of Piagetian theory and its history of dissemination is almost unique in the literature on Piaget for its efforts to relate the theory to broader issues of scientific methodology. Groen's discussion of the implications for education of Piagetian theory should prove valuable to a variety of readers.

In Chapter 5, Hermine Warren and I examine the influence of psychoanalysis in American elementary education during the period 1900 to 1957. I believe that our chapter is the first detailed account of these matters and for this reason alone represents an interesting case study of the impact of one of the most important psychological theories of this century on American education.

In Chapter 6, H. H. Stern, Marjorie Bingham Wesche, and Birgit Harley analyze the impact of the language sciences on second-language education. They examine in considerable detail the developments in linguistics, psycholinguistics, and social linguistics over the past thirty years that have affected the teaching of second languages. They make a special effort to identify the impact of linguistic research on an extensive set of practical problems, ranging from the preparation of curriculum materials to the attitude of teachers.

In Chapter 7, J. W. Getzels examines a number of cases and a variety of issues that have been widely discussed on the relation of basic research to educational practice. One of his most interesting examples, quite different from others discussed in this volume, concerns the use of theory-oriented work in the architectural design of classrooms.

In Chapter 8, Torsten Husén analyzes in considerable historical detail the relation between educational research and educational reform in Sweden since 1945. He writes with the authority of one who has played a prominent role in these matters throughout the entire period.

In Chapter 9, Marvin Bressler and Joel S. Berke examine the relation between social science research and the law, especially as reflected in the *Rodriguez* case, and the issue of school-finance reform. As they emphasize, judicial resolution of educational issues has been a prominent feature of education in the United States in the last three decades, and the use of research results in the judicial decision-making process has been evident.

Although these nine studies by no means exhaust the range of instances in which research has influenced educational practice, they do provide a reasonable sample of some of the most prominent instances of impact. It is easy to recognize how different they are in character from the eight cases studied in the Battelle Report, but the existence of these obvious differences in no way implies a judgment as to which sort of contribution is more important to our society. The obvious answer is that both sorts are required and both play a necessary role.

Educational research as a discipline scarcely existed prior to the 20th century. Some of the chapters are devoted to topics that are bound to be remembered historically as among the most important examples of educational research in this century, but I like to think of the case studies presented here as representing only a sampling of the beginning of educational research in its early years of development. What the future holds should be brighter and better, because it will be able to build on the kind of work reported in this volume and, as that building takes place, the impact on practice should become more marked.

The framework for the work done on these nine case studies was provided by support from a contract (NIE-C-74-0116) from the National Institute of Education, U.S. Department of Health, Education, and Welfare. Arthur Melmed of the National Institute of Education has been helpful at all stages, from initial conception to completion of the printing of the present volume. Without his continued help and support, the present volume would not have come into existence.

This project was almost completed during my term of office as president of the National Academy of Education, which ended in October 1977. I cannot express too strongly my appreciation of the dedication of the staff of the Academy during my term in office, who helped extensively on this project. I mention especially William D. Hyde, Jr., who was executive secretary of the Academy from January 1975 to June 1977, Dianne Kanerva, who has had primary editorial responsibility for the volume, Pentti Kanerva, who has had main responsibility for the fast and efficient computer-based typesetting, and Bonnie Senko and Marguerite Shaw, who have both helped in innumerable ways in preparing the volume for publication.

Patrick Suppes

Stanford, California
December 1977

Chapter 1

ON THE THEORY-PRACTICE INTERFACE IN THE MEASUREMENT OF INTELLECTUAL ABILITIES

by

JOHN B. CARROLL
*University of North Carolina
at Chapel Hill*

A PRIME goal of scientific discovery and technological development is the production of new knowledge—knowledge that can be disseminated through communications in scientific journals, monographs, books, and the like. This knowledge can be taken advantage of, generally, only at the cost of some special attention and study on the part of the user. Much of the output of psychological and educational research is of this character; that is, the fruits of this research are mainly writings (like the present volume) that are intended to provide educators with guidance as to how various aspects of schooling should be conducted.

Often, however, particularly in the physical and biological sciences, scientific knowledge can become embodied in some kind of physical object or substance—a “long-lasting” razor blade, a color television set, or a new drug or vaccine. Most of the time, a razor blade, a color television set, or an over-the-counter drug can be employed quite successfully by the ordinary layman who has not the faintest idea of the technology in metallurgy, electronics, biochemistry, or whatever, that went into its development and production. There are, of course, many technological products, such as supersonic airplanes, electron microscopes, or radiation therapy machines, whose use requires a high degree of specialized training.

At least some of the products of educational research are concrete physical things. One of these products—insofar as it may consist of a set of one or more physical objects like test stimulus arrays, response devices, or scoring keys—is the mental test. The mental test is a concrete product of the subdiscipline and technology known as psychometrics that is some-

what analogous to the knowledge-embodiment products of technology in the physical and biological sciences.

As in the case of a new drug, it would not ordinarily take much skill for a person without specialized training to administer the test to himself or to others. Without a knowledge of the circumstances it would be difficult to predict whether such an action would have beneficial or adverse consequences. Certain it is, however, that a layman (or, particularly, a person who undergoes a "psychological test") would ordinarily be unaware of the technology that produced the test, just as he would be unaware of the technology involved in the discovery, development, and production of the drug. And just as a layman might be inclined to have reasonable doubts about the usefulness or efficacy of a drug he is instructed to take, so a layman (or a test subject) might entertain reasonable doubts as to the value and precision of a psychological test. A mental test usually *looks* like a rather simple object, consisting, perhaps, of a series of tasks based on visual or auditorily presented verbal, mathematical, or symbolic materials. The scientific knowledge that may have gone into the selection and construction of these tasks, or that would be employed in the analysis of performances on these tasks, would hardly be evident from superficial examination of a mental test.

The intention of this chapter is to give a view of the development and present status of the scientific knowledge that lies at the base of psychological tests, with special reference to tests of "intelligence" or intellectual capacity. It suggests that no matter how simple (or apparently simpleminded) the test content may appear to be, the selection or devising of this content, and the analysis of people's performances on these tests, is undergirded with a substantial and respectable amount of knowledge and technology and that although the present theory and technology are highly sophisticated, there are reasonable grounds for questioning at least some of it and for thinking that new and profitable directions for research can be found.

The focus is on tests of so-called intelligence, IQ, or mental ability, partly because of space limitations, but more because this is currently (and has been since its beginnings) one of the most controversial fields of testing. Psychological and educational tests exist in great variety and profusion, as can be verified by consulting O. K. Buros's latest compilation, *Tests in Print II* (1974). Much of the basic psychometric technology to be discussed here pertains to all types of tests. Problems of reliability, validity, scaling, standardization, and the like have to be considered in connection not only with tests of mental abilities but also with tests of educational achievement, personality, interests, attitudes, motor skills, and the rest. But there are somewhat special considerations—linked to basic psychological issues about the nature of human capacities—that apply with particular force in the domain of tests of intellectual abilities; these

will be discussed here in due course. Little attention will be paid to the measurement of educational achievement except insofar as tests of intellectual ability may be said to be, or can be shown to be, tests of educational attainment.

Even in considering the field of mental ability testing, we must limit our scope considerably. Except to outline historical antecedents in other countries (particularly England, France, and Germany), the focus will be on developments in the United States. Little consideration will be given to the use of IQ tests in clinical practice to diagnose mental retardation and so-called learning disabilities. The emphasis will be on "group" tests—that is, tests given to fairly large groups of people at a time and usually of the "paper-and-pencil" variety. Starting roughly in the 1920s, such tests have often been used in schools to yield measures of intellectual capacity to serve as a partial basis for ability grouping, selection for special classes, etc., and tests of this type have increasingly been used in connection with admission to higher education.

The approach will be descriptive and historical. We will trace the history of the mental testing movement, with special attention to its technological aspects, through two somewhat arbitrarily defined periods: (a) an "early" or "developmental" period beginning in the late 19th century, during which the foundations of theory and practice in psychometrics were laid down; and (b) a "modern" period starting around 1935 with the founding of the Psychometric Society and its journal *Psychometrika*, during which many refinements were made in the technology of testing and during which testing became a major enterprise.

Each period will be introduced with an overview of its major developments and achievements. This introduction will be followed by a detailed discussion of particular topics that are regarded as essential to the scientific study of tests of intellectual ability and to the orderly development of their applications in educational settings. These topics, which are closely interrelated, are as follows:

1. *Theories and doctrines concerning the nature of what is to be measured.* All measurement has to do with (at least) the ordering (placing in an ordinal series) of a defined class of objects with respect to some attribute or attributes. In the case of intellectual ability, we are seeking to provide ways of ordering individuals with respect to certain attributes of capacity or behavior. The question considered here will be: Exactly what attributes are involved when we try to develop procedures for ordering or measuring individuals with respect to "intellectual ability"? More fundamentally, to what extent is it reasonable to think of intellectual ability in terms of traits or attributes?

2. *Theories of measurement and scaling.* In each historical period, we must consider the question: What basic theories or doctrines about measurement and quantification have guided the development of psychometric

procedures?

3. *Statistical theory and computational technologies.* Insofar as psychological measurement involves quantification and the study of distributions of observations, it relies heavily on general statistical theory. Also, in view of the (often) large numbers of such observations that may be involved in either research or the practical applications of psychological measurements, advances in computational technology have been important in the analysis and use of the data so obtained. For each historical period, pertinent developments in statistical theory and computational technology will be reviewed.

4. *Theories of psychological measurement.* Considered here are the developments under what is ordinarily called "mental test theory," that is, the special formulations of measurement theory that have been regarded as necessary in connection with "fallible" measurements, especially with regard to the "scores" derived from series of (usually binary, or "pass-fail") observations of task performances.

5. *Procedures of test construction, standardization, and application.* Under this topic we consider the implications of the preceding four topics for specific procedures of test construction, standardization, validation, and use.

6. *Practices in the use of tests and research on such uses.* Although the people and organizations that construct tests usually make efforts to spell out how they believe these tests should be applied in practice, their actual use in school settings may not conform to the injunctions contained in test manuals and the like. Under this topic we consider research concerned with how tests have been, or are, actually applied in practice.

Following the discussion of each of these topics for the historical period in question, a general appraisal of the developments of the period will be given from the standpoint of what impact research and theory in psychological measurement had upon the general conduct of education.

1. THE PERIOD OF DEVELOPMENT

1.1 Overview

For purposes of the present account, I have considered as a single continuous period the span of time that began with the publication of Sir Francis Galton's (1869) early work, *Hereditary Genius*, and that ended with the founding, around 1935, of the Psychometric Society and its journal *Psychometrika*. This would appear to be a long period, but it is not really very long; its duration—66 years—is about that of the average life span. (Indeed, some of the more prominent figures of this period are still living or have died only recently.) Yet, during this period, the science of mental testing progressed from its crudest beginnings to a stage of high development, leaving only matters of refinement for the period from 1935

to the present. It was during this earlier period that the major problems in the study of mental abilities were identified, the basic methodologies were developed, and tests of mental ability came into wide use. More detailed accounts of this period, or of parts of it, are available in works by Peterson (1925), Linden and Linden (1968), and DuBois (1970).

It might be interesting to trace the antecedents of the theory of individual differences in mental ability, and the methods of assessing these differences, in ancient, medieval, and early modern times; but there does not seem to be available any thoroughgoing study of these antecedents. From the works just cited, which include brief accounts of theories and trends prior to 1869, one gets the impression that the concept of individual differences in mental ability was very slow to develop, and the procedures for assessing such differences were very crude from the standpoint of current technology. Since the focus of this essay is on the development of our present theories and methodologies, I shall not attempt any treatment of the early historical antecedents.

Mental testing developed within an intellectual, societal, and educational context that included, for example, Darwin's formulation of the theory of evolution and the consequent movement described by Hofstadter (1944) as social Darwinism, the later phases of the Industrial Revolution, the marked increase in the population (particularly in the United States through immigration), and the increasing democratization of the schools, with the Progressive Education Movement (Cremin, 1961). Social Darwinism, in England and elsewhere, focused attention on differences, presumably of a hereditary nature, in the adaptability of human individuals to the new demands of an industrialized and technological society. Increase in the population led to the development of programs of education for the masses, and the democratization of the schools demanded the creation of ways of handling schooling for young people who would not have been admitted to the formerly highly selective school systems. In addition, the period was marked by a major war, which involved problems of selecting and training millions of individuals for combat. As it happened, the mental testing movement had progressed far enough, by the time of World War I, to be ripe for trial in a mass selection program that engaged the efforts of some of the most capable and prestigious psychologists in America.

Like many others in the period before the turn of the century, Galton believed that mental ability could be revealed by performances on rather simple tests of the acuity of the senses, the speed of reaction, and the like. Even early in the 19th century it had been recognized, by the astronomer Bessel, that individuals differed reliably in speed of reaction; also, a primary preoccupation of the early psychologists, like Fechner (1860) and Wundt (1862), had been with the measurement of powers of sensory discrimination, resulting in the theory and methodology of psychophysics,

the science of quantitative relations between physical magnitudes and sensations.

In his *Hereditary Genius* (1869), Galton drew attention to the problem of delineating the characteristics of persons of genius, suggesting that genius is a trait that tends to run in families, but it was not until the publication of *Inquiries into Human Faculty and Its Development* (1883) that his work on the development of tests of mental abilities came to public attention. Galton can, in fact, be credited with the invention of the mental test, although it was only several years later that Cattell (1890) introduced the term *mental test*. From 1884 to 1890, Galton collected data from the general public on a variety of simple mental tests at the Anthropometric Laboratory that he established in London. Although it was about this time that he also developed the notion of correlation, Galton did not derive any important conclusions about individual differences from this work, except possibly the fact that these differences tended to be distributed according to the Gaussian law of error.

Working with Wundt at Leipzig, the American psychologist James McKeen Cattell became interested in individual differences in speeds of reaction, sensory discrimination, word association, and other simple mental tasks (Cattell, 1885, 1886a, 1886b). Returning to America as professor of psychology at the University of Pennsylvania, Cattell set about applying such tests in his laboratory; shortly afterwards, taking a position at Columbia University, Cattell postulated that these tests might be predictive of the scholastic success of college students (Cattell & Farrand, 1896). Unfortunately, Cattell's hopes were dashed when an analysis of his data by Wissler (1901)—reportedly the first substantial application of the correlational method—showed that the scores of the tests that Cattell had given several years earlier to Columbia University undergraduates had no practically useful correlations with their college grades. The results of this study, as well as those of a somewhat similar one by Stella E. Sharp (1898–1899) working under Titchener at Cornell, were taken as persuasive evidence that mental tests of simple mental reactions had no promise as predictors of scholastic achievement nor, for that matter, as measures of anything like intelligence.

From the debates in the literature of the times, one would think that the mental-testing movement was being laid to rest. There were, however, activities elsewhere that presaged a very healthy development.

The German psychologist Ebbinghaus (1897), already well known for his studies of memory, devised a "completion" test for measuring the effects of fatigue on children's school performances. Ebbinghaus's test very much resembled what would be today called a "cloze" test (Taylor, 1953); it required the child, using a guessing procedure that depended on comprehension of the total context, to fill in blanks representing deleted syllables or words in a prose passage. Ebbinghaus noted the substantial

correlation of scores on this test with children's performance in school and with teachers' estimates of brightness.

The French psychologist Binet, however, may be credited with discovering an appropriate way to measure intelligence or mental ability. After a long period of investigating the simpler tasks studied by Galton, Cattell, and others, Binet came to the conclusion that tests comprising more "complex" tasks, resembling the mental activities of everyday life, were more promising. The story has often been told of how he developed "*une échelle métrique d'intelligence*" to distinguish mentally retarded children from children of normal intelligence and school progress. The scale (Binet & Simon, 1903) consisted of a series of tasks of increasing difficulty, each task representing the typical performance of children at a particular chronological age level. The tasks were highly varied, but most of them relied in some way on the understanding of language and the ability to reason with either verbal or nonverbal (spatial, numerical) materials. This scale, and the more refined one published shortly afterwards (Binet & Simon, 1908), became the basis for individual intelligence tests in a number of countries and languages. In fact, despite a number of revisions (Goddard, 1908; Terman, 1916; Terman & Merrill, 1937, 1960), the content of the present "Stanford-Binet" scale can in most cases be traced back to that of the original scale.

Ebbinghaus unwittingly demonstrated the feasibility of group tests of intelligence, and it was early recognized by a number of American psychologists (Whipple, 1910; Otis, 1918 [though the actual development occurred several years earlier]) that many of the kinds of mental tasks included in the Binet-Simon scale could be adapted for group testing. As will be pointed out in our discussion of the psychometrics of group testing, there was a subtle, perhaps radical, change in the method of deriving scores from group tests. Whereas the Binet scale called upon the examiner to find that point or region on the scale of task difficulty that separated generally successful performance of a child from generally unsuccessful performance, the group testing method involved simply asking the examinee to try as many "items" as he could—usually within a set time limit—and the score was computed simply as the number of items correctly answered. The resulting scores were regarded as being measured on a "point scale."

There was another subtle change entailed in many of the early group tests. Whereas Binet-type items usually called for a definite answer supplied by the child (e.g., a definition of a word, an answer to a numerical reasoning problem, or the like), group tests tended to call merely for recognition of a correct answer among several possible options. Exactly when the "objective" or "multiple-choice" form of item was first used is not clear, but this format quickly became characteristic of nearly all group tests of mental ability.

Group testing contrasted with Binet-type individual tests also in the fact that testing was done almost exclusively by means of "paper-and-pencil" printed tests. This often required that the examinee be able to read; it seems that test constructors blithely assumed that the examinees would all be equal in their ability to read and that only their ability to handle the material presented in printed form was at issue.

This, in fact, was apparently the assumption of the group of prominent American psychologists who were called together to devise ways of giving psychological examinations to Army recruits at America's entry into World War I. In an amazingly short period of time, and yet after several tryouts and revisions, a test called the Army Alpha Examination (in several alternate forms) was produced and administered to more than a million Army men. Based on previous work of such investigators as Goddard, Otis, Bingham, Wells, and Thurstone, it consisted of eight brief subtests that could, in all, be administered in about 25 minutes. Early trials revealed a surprisingly high rate of illiteracy in the recruit population, a finding that resulted in a separate effort to develop a test, Group Examination Beta, appropriate for illiterates and recruits who did not know English. Scores on both of these tests were scaled in terms of mental age, the concept originated by Binet. The concept of the IQ (intelligence quotient) originated by Stern (1912) for assessing children's rate of mental growth was (correctly) regarded as inappropriate for use with this adult population.

The wide publicity given to group testing during World War I was attended with keen interest by educators and school administrators, who felt that group mental tests held great promise in helping teachers assess the learning capacities of their students. There ensued a flood of tests that imitated the Army tests in content and format. Many, it appeared, were constructed by individuals who paid little attention even to the relatively primitive technicalities of test construction that had been developed in the Army testing program. Others were developed by some of the psychologists (Otis, E. L. Thorndike, Thurstone, etc.) who had been involved in the Army work. Notable examples of such tests were the *National Intelligence Test* (designed for Grades 3-8), the *Terman Group Test of Mental Ability* (designed for Grades 7-12), and *Thurstone's Psychological Examination for College Freshmen*. Notions of "standardization" (the derivation of norms for representative samples of the population for which a test was designed) and "validation" (establishment of substantial correlation of test scores with independent measures of the attributes being tested or predicted) were developed during this period. Some of the tests recognized somewhat independent dimensions of ability; for example, beginning in about 1925, Thurstone's *American Council on Education Psychological Examination for College Freshmen* yielded two scores, a "linguistic" score and a "quantitative" score; similar scores ("verbal" and

"mathematical") were yielded by the Scholastic Aptitude Examinations constructed by C. C. Brigham for the College Entrance Examination Board, starting in 1926. The culmination of this trend, for the period under consideration here, was represented by E. L. Thorndike's development in 1927 of his CAVD test, which contained four quite different parts (Completion, Arithmetic, Vocabulary, and Directions) and yielded cores on "altitude," "width," and "area."

These and many other tests became widely used in schools and colleges, and the publication and distribution of tests grew into a large, and often highly lucrative, enterprise. The concept of "intelligence" or "scholastic aptitude" came into general acceptance, but the notion that intelligence or scholastic aptitude reflected largely the effects of native endowment in interaction with schooling was, apparently, slow in coming. Writing in 1925, Joseph Peterson had this to say:

The popular mind today . . . neglects innate individual differences. Many of the weaknesses of our present educational system are directly due to the view, now disappearing, that we are born "free and equal" in the sense that all alike can learn what the schools teach. Failures, according to this view, are attributed to stubbornness, laziness, etc., rather than to inability or to innate differences in the structure and organization of individuals. (p. 15)

This sketchy overview of the development of group tests of mental ability has omitted many details, and even some major developments. Some of these details will be filled in as we consider specific aspects of the testing movement as it developed during the period now in view.

1.2 Notions of Mental Ability

What does one seek to measure in developing a test of "mental ability"? One can interpret the efforts of Galton, Cattell, Binet, and many others in the developmental period of mental testing as oriented toward the diagnosis and verification of an "ordinary language" concept of intelligence whereby it is recognized that human beings exhibit, somewhat independently of the amount of education to which they have been exposed, grades of intelligence ranging from idiocy or feeble-mindedness, through average, up to genius. Galton was most interested in the characteristics of genius; Binet, at least in his earlier practical work, was concerned with distinguishing grades of mental deficiency and contrasting them with the abilities of the average child. Galton was unsuccessful in discovering tests that would exhibit the mental powers of a genius; he must have been surprised to find that recognized geniuses had no greater prowess in sensory acuity and speed of reaction than the average man. By focusing his attention on relatively complex mental tasks, Binet was successful in devising a scale that distinguished grades of mental deficiency. The psychologists who developed the Army Alpha and Beta examinations were evidently attempting to take Galton's advice seriously, namely, that

one should "obtain a general knowledge of the capacities of a man by sinking shafts, as it were, at a few critical points" (Galton, in a letter appended to Cattell, 1890). For these Army psychologists, the critical points were abilities to understand language, to perform reasoning with semantic and quantitative relationships, to make "practical judgments," to infer rules and regularities from data, and to recall general information. Operationally, every one of the more successful tests of intelligence, whether of an "individual" or "group" type, employed a rather wide array and variety of tasks involving the understanding and manipulation of verbal and nonverbal materials and problems. While some of these tasks intersected those specifically taught in school, the emphasis was on tasks that reflected the individual's ability to profit from his total experiential history.

The general view of the nature of intelligence implied here was later expounded in considerable detail by E. L. Thorndike (Thorndike, Bregman, Cobb, & Woodyard, 1927). Thorndike saw intelligence as a general capacity that manifests itself in a large variety of tasks; it is, in effect, a capacity for forming bonds or connections among ideas, concepts, etc. Individuals of high intelligence are those who have the capacity to form a large number of bonds and have had the opportunity (through experience, education, etc.) to do so. Insofar as the capacity to form bonds might be regarded as innate and the actual formation of these bonds is thought to be a result of appropriate opportunities to form them, Thorndike viewed intelligence as having both hereditary and environmental components.

A somewhat similar theory, the "sampling theory" of intelligence, was proposed by a Scottish psychologist (and chief author of a series of "Moray House" intelligence tests used widely in schools in the United Kingdom), Godfrey Thomson (Brown & Thomson, 1921). According to this view, any mental task "samples" a wide variety of mental operations; the correlation between two tests is therefore a function of the amount of overlap between the sets of operations thus sampled. (In a sense, Thomson foreshadowed the "information processing" approach which is in favor in contemporary psychology; Thomson, however, never attempted to specify the details of how mental operations enter into tasks.)

Nevertheless, a satisfactory *definition* of intelligence or mental ability was always elusive. The few definitions that were enunciated by leaders in mental testing seldom corresponded very well with the actual measurement procedures embodied in their tests. For example, one definition of intelligence offered by Binet (1890) would hardly yield suggestions as to how to operationalize the concept:

What we call intelligence, in the narrow sense of the term, consists of two chief processes: First, to perceive the external world, and then to reinstate the

perceptions in memory, to rework them, and to think about them. (p. 582, translated from the French)

At another point, Binet (Binet & Simon, 1909, 128ff.) laid down three criteria for intelligent thought: (a) *la direction*, the taking and maintaining of a given mental set; (b) *l'adaptation*, the adaptation of thought for the purpose of obtaining a given end; and (c) *la critique*, the taking of a critical attitude toward one's thought, and correcting it where necessary.

We have seen that, in the early years of the period we are considering, use was made of relatively simple tasks, usually involving powers of sensory acuity and judgment (e.g., detecting small differences in the weights of two visually similar objects) or speed of reaction time in responding to stimuli (e.g., naming colors). Selection of these tests was based on the apparently plausible hypothesis that intelligence resides in the power of an individual to respond sensitively and quickly. Of these early types of tests, perhaps only one has survived in current measures of intelligence—the memory span test—and even this is recognized to be a measure of a rather special dimension of intelligence. It was only when Binet developed tests of more “complex” processes that he could feel that he was measuring something that revealed mental ability differences. It may be noted, however, that Binet's approach was *developmental*: He was essentially the first to trace the development of certain abilities over the age range of childhood and early adolescence. A criterion for the acceptance of a task into his series was that it show not only progressive increase in performance over these ages but also consistent relationships with degrees of mental deficiency over these ages. (Otherwise, a test of a motor skill like walking might have been included.)

A symposium on the meaning of intelligence published in 1921 (Thorndike et al., 1921) produced a tremendous profusion of definitions and opinions. Intelligence was variously described as “ability to learn” (Buckingham), as “the power of good responses from the point of view of truth or fact” (Thorndike), as “the ability to carry on abstract thinking” (Terman), as “the ability of the individual to adapt himself adequately to relatively new situations in life” (Pintner), as “involving two factors—the capacity for knowledge and the knowledge possessed” (Henmon), or as “the capacity to acquire capacity” (Woodrow). Beardsley Rumel, who later, as an economist, introduced the idea of the withholding tax, declined to enter the debate because there was insufficient precision in the terms and concepts to form a basis for discussion and because of “an absence of factual material on so many essential points.” Henmon added the caution that “the so-called general intelligence tests are not general intelligence tests at all but tests of the special intelligence upon which the school puts a premium.”

Perhaps it was truly premature to expect any consensus on the mean-

ing of intelligence at this early date.¹ Yet the tests must have been measuring *something*, and to a degree their reliability as measurements, and their validity in predicting school and other kinds of performances, had been established. Despite the fact that Binet's definitions of intelligence differed considerably from those espoused by the authors of group tests of intelligence, there was evidence from the work with the Army that scores from the Army Alpha Examination correlated rather highly with ratings derived from tests of the Binet type (Yerkes, 1921).

Wholly unmentioned in our overview of this developmental period was the work of the British psychologist Spearman, who very early in the century published a strikingly original interpretation of some data² that he had collected in a village school in Hampshire, England (Spearman, 1904b). Setting up a matrix of correlations among the test scores and academic ranks, Spearman noticed that they could be arranged and analyzed in a special way, "hierarchically," to show that all the variables measured just one "factor" in common, but to different degrees. From this, Spearman developed what he called (somewhat misleadingly) the "two-factor" theory of intelligence, whereby each test of a set is regarded as measuring one "general" factor in common with all the other tests and, in addition, a "specific" factor that is unique to that test. Spearman devoted the major part of his professional life to attempting to establish the universality of this two-factor theory and, in addition, to explicating the psychological nature of the general (*G* or *g*) factor. The major writings of Spearman that bear on this subject are the books *The Nature of 'Intelligence' and the Principles of Cognition* (1923), *The Abilities of Man* (1927), and (with L. Wynn Jones) *Human Ability* (1950). This work embroiled Spearman in many controversies, for he was often accused by his critics of "selecting" his data—that is, selecting tests in such a way as to cause the data to fit the two-factor theory. Eventually, Spearman conceded that not all tests could be interpreted solely in terms of a two-factor theory; his later writings recognize the appearance of "group" factors (and even two other specialized "general factors"—perseveration and oscillation) alongside the original *g* factor. What Spearman's original *g* factor seems to correspond to, in current interpretations, is the fact that almost any set of mental tests, no matter how different, will *tend* to exhibit *positive* intercorrelations (assuming that the test scores are all oriented in the same way, i.e., that numerically "high" scores represent the more desirable, compe-

¹ Two years later, the psychologist E. G. Boring (1923), in an article in the *New Republic*, made the not entirely frivolous suggestion that "intelligence is what the tests test"—an admittedly "narrow definition, but a point of departure for a rigorous discussion . . . until further scientific discussion allows us to extend [it]."

² The data, surprisingly enough to the modern reader, consisted of scores on tests of discriminative judgments of pitch, light, and weight, along with marks in several academic subjects. Spearman chose these tests after an extensive review of the work of Cattell and others who had attempted to use sensory judgments in measuring intelligence.

tent behaviors). What really interested Spearman was the psychological nature and interpretation of whatever it is that tends to produce positive correlations among all mental tests. To this end, he held that the "general" factor is central and supreme in all tests of intelligence. He even thought it had identifiable physiological correlates.

Spearman's *Nature of 'Intelligence' and the Principles of Cognition* (1923) may be thought of as the result of a detour he took out of statistical psychology into cognitive psychology. Virtually unknown to today's cognitive psychologists, but potentially highly rewarding for them if they would attend to it, it was a thoroughgoing analysis of reasoning processes as they are exhibited in solving syllogisms, making inferences from propositions, solving mathematical problems, and the like. From this work, and from the empirical data (much of it from group intelligence tests) amassed in *The Abilities of Man* (1927), he arrived at the notion that the "general" factor represents the power of reasoning, or "the neogenesis of abstract entities." To the extent that a test measures a general factor, Spearman saw in it operations of the "eduction of relations" and the "eduction of correlates." A vocabulary test, for example, often requires the subject to notice similarities or differences in meaning, while a verbal analogies test ("Father is to mother as son is to _____") requires the noticing of correlative or analogous relationships. On this theory, an intelligence test (to the extent that it is "loaded" with *g*) measures the extent to which an individual can perform such mental operations.

At about the same time that Spearman published his *Nature of 'Intelligence'* Thurstone published a book with a similar title (1924), but it was concerned more with the nature of perceptual processes in intelligence and for various reasons failed to have the impact that Spearman's work did. Indeed, it was all but ignored by Thurstone himself in his later work.

Out of the controversies surrounding Spearman's two-factor theory and the attempts to put it to empirical test, there gradually emerged a somewhat different, and perhaps better articulated, view of mental abilities. Spearman himself, but even more his co-workers and students, sought to explain the fact that matrices of correlations among test variables did not always conform well to the two-factor theory. It was noticed that specific groups of intelligence tests—by virtue of special similarities in content, format, or the response processes involved—tended to exhibit intercorrelations that were greater than would be predicted by the two-factor theory. Already in 1927, Spearman recognized the existence of group factors of intelligence apart from *g*. In cooperation with the American educational psychologist Karl Holzinger, Spearman and his associates initiated a "Unitary Trait Study" in the early 1930s to seek information on what kinds of "group factors" could be established empirically and interpreted psychologically in useful ways.

In the meantime, Thurstone, who had early been involved in the Army

examining program and who had been one of the most creative test developers, was seeking improved mathematical formulations of the problem of identifying "factors" of intelligence. The story of these efforts by Spearman, Holzinger, Thurstone, and others to delineate the factors of intelligence belongs more properly, however, to the discussion of the "modern period."

1.3 *Theories of Measurement and Scaling*

Psychological measurement involves difficult problems concerning the nature of the scales on which measurements are taken, the units of measurement, the existence of an origin or zero point, and the numerical operations that can legitimately be performed with the measurements. No indications have come to my attention that early developers of intelligence tests concerned themselves with such problems. To be sure, in his *Grammar of Science* (1892), Karl Pearson (who will be mentioned in a later section as a collaborator of Galton in the development of the correlation coefficient) recognized the need for a science or theory of measurement, but he did not discuss any peculiarly psychological problems of measurement. Galton, Spearman, Binet, and other early workers in the measurement of intelligence apparently relied on traditional concepts of measurement as they had come down from Euclid and Newton. Modern doctrines of the foundations of measurement took shape only in the work of the philosopher-physicist H. von Helmholtz (1885) and the mathematician Hölder (1901), but this work was not to influence psychological measurement until much later, for example, in the writings of B. O. Smith (1938) and S. S. Stevens (1951).

It is, nevertheless, instructive to examine the assumptions about measurement that were made by early developers of intelligence tests. It is convenient to depart somewhat from a strictly chronological treatment in order to begin with a look at the rather special measurement characteristics of Binet's mental age scale. Subsequently, attention will be given to the tradition that began with Galton and other British psychologists and that was carried on by group intelligence testers in America.

Implicitly, the first problem that confronted Binet was whether mental ability varies continuously (i.e., exists in different gradations that can be as small as one likes) or exists in some set of discrete steps or classifications. He was, after all, set the task of distinguishing "abnormal" or mentally retarded children (as a group) from "normal" children (as a group). As a result of his empirical work, he apparently came to realize quite early that mental ability or intelligence can best be thought of as varying continuously and that different degrees of intelligence can be ranged along a scale. (In Binet's time, the dominant theories of individual differences, especially in personality, were essentially typologies. Thus, Binet's achievement in recognizing the continuous variation of intelligence was

perhaps more significant than it seems today.)

The following quotation from Binet suggests still other implications about measurement theory.

The basic idea of this method is the establishment of what we shall call a metric scale of intelligence [*échelle métrique de l'intelligence*]; this scale is composed of a series of tests, of increasing difficulty, starting at the lowest intellectual level that one can observe, and ending at an average and normal level of intelligence, wherein to each test there corresponds a different mental level. (Binet & Simon, 1905, p. 194, translated from the French)

What is striking here is the idea of a scale composed of tasks of increasing difficulty where, in addition, each task corresponds to a different level of ability or "mental age" (a concept Binet originated). There was nothing like this in the previous history of mental testing, and, in fact, there are few parallels to such a scale in ordinary experience (except perhaps the erection of bars at increasing heights to test jumping ability, or the testing of minerals for hardness by finding which ones of a graded series of substances—from soapstone to diamond—could produce scratches in them). One other idea that is implicit here was perhaps not so new—the idea that the outcome of each task would be binary, that is, would consist of either passing or failing.

Binet was, of course, a little too optimistic in seeming to suggest that each task would correspond exactly and reliably to a particular mental level. As the Binet scale developed, it was found necessary to use a number of tasks at each level of "mental age." From this point on, the task of determining a person's mental age was reminiscent of one of the psychophysical methods devised by Fechner, Wundt, and others to determine the level of a person's sensitivity to faint stimuli or to a small physical difference in stimuli. In fact, it is in the psychometric analysis of a Binet-type scale that we see the strong resemblance of psychometric methods to those of psychophysics, a resemblance that was capitalized on in later developments in "test theory" (of which more will be said below). That is, just as one tests hearing ability by presenting sounds of increasing loudness or softness, and determines the limen (threshold) at which the individual hears a sound half the time, so one determines mental age by presenting different tasks in order to find the level on the mental age scale where the individual can pass half (or some other fraction) of the tasks. Such testing is readily done when the examination is conducted for one child at a time; it is much more difficult to accomplish under group testing conditions.

Later developments of Binet-type tests, particularly at the hands of Terman (1916) in his "Stanford-Binet" test, led to an elaborate technology for this type of scale, involving the careful selection and placement of tasks on the scale in such a way as to produce approximately constant means and standard deviations of intelligence quotients over different

ages. (This implied, of course, that the scale was conceived not only to apply to normal and mentally retarded children but also to the upper ranges of ability.) The idea of an intelligence quotient came from the German psychologist Stern (1912), who had noticed that as chronological age increased, variation in mental ages increased proportionally; thus, he saw that it would be possible to obtain a ratio whose standard deviation would be approximately constant over chronological ages if mental age was divided by chronological age.

Despite the interesting psychometric characteristics of the Binet-type scale, neither Binet nor any of his followers gave much thought to the question of whether such a scale would be homogeneous in the sense that every task or item measures the "same" kind of mental ability. One can perceive that this would not necessarily be the case. Binet's scale depended mainly on the fact that nearly all human functions, whether cognitive or not, tend to mature at some specified rate, and more or less in a fixed order. But different mental functions could mature at different rates. Intelligence as derived from Binet-type tests is therefore essentially a measure of average rate of growth of various mental functions—not a measure of *momentary* mental growth (in the sense of a first derivative), but of rate of mental growth as estimated from the cumulative performance or achievement up to the time of testing. It can easily be shown that as a consequence it has an element of constancy automatically built into it, if one assumes that competencies once achieved are not forgotten or lost. Amid the discussions of the "constancy of the IQ," the implications of this fact were not clearly seen until a later phase of the history of mental tests.

Let us turn now to the measurement assumptions underlying the group testing tradition. The kinds of tests employed by Galton and Cattell in the early "experimental laboratory" phase of the testing movement most often yielded scores based on direct physical measurements—speeds of reaction expressed in fractions of a second or the physical magnitudes of stimuli or stimulus differences that could be apprehended by the subject. As such, they had the properties of other kinds of physical measurement. Very few of these tests, therefore, had "binary score" characteristics. It was assumed, presumably, that any mental abilities measured by these tests were in direct relation to the score levels attained—an assumption that probably was justified.

When Binet-type items began to be applied in group tests of intelligence, shortly after the turn of the century, the measurement assumptions changed rather radically, although it is probable that few psychologists gave much thought to this change. A "test" became an assemblage of items; generally each item or task yielded a binary score (pass or fail). The test score became the number of such items that were passed and the resulting scale was called a "point scale." When the test was

given under a time limit, the score became the number of items that the examinee was able to attempt *and pass* within that time limit. Test constructors recognized that the items might differ in difficulty; in fact, the usual practice (as it still is today) was to arrange the items in approximate order of difficulty, starting with the easier ones, apparently under the assumption that this made for a better adjustment of the examinee to the demands of the task as a whole. Also, it meant that the slower examinees were likely to attempt a greater number of items than otherwise, if difficulty of an item entailed a greater amount of time. As in the case of Galton's and Cattell's tests, it was assumed that whatever mental ability was being measured by the test would be in direct monotonic relation to the total score, that is, to the number of items passed. Such an assumption was somewhat more questionable in this case than in the case of the simple tests employed by Galton and Cattell, for several reasons. First, there was an inadequate guarantee that the separate tasks were homogeneous with respect to what they measured; two persons might obtain precisely the same score yet pass completely different items and thus exhibit quite different mental abilities. Imagine an analogous situation of a test with half the items in English, half in French; if a person scored 50% correct, it might be that he passed only the items in English, or only the items in French, depending on which language he knew. In the group intelligence tests that were constructed in the early days, the only guarantee of item or task homogeneity was the judgment of the test constructor. Second, scores could depend upon either the level of the examinee's abilities (which could be thought of as the probability that he could pass any item picked at random) or the speed with which he could attempt them, or upon some combination of the two. Third, the difficulty levels of items were assessed in terms of the average performance of a group and would not necessarily correspond to their difficulty levels for a given individual.

The seriousness of these problems of scaling was seldom perceived in the developmental phases of the mental testing movement. We find only hints of it in the writings of E. L. Thorndike. Perhaps the earliest clear recognition of the relation between the patterns of scores on individual items and individuals' total scores is to be found in the work of David A. Walker (1931, 1936, 1940), a Scottish psychologist working with educational achievement measurements. Walker pointed out that this relation could range from what he called "hig" (from "higgledy-piggledyness") to its converse, "unig." The "unig" situation would exist when "a given answer-pattern completely determines the score-scatter," that is, each score x is composed of correct answers to the x easiest items, whereas maximum "hig" would occur when all the scores with value x might be composed of a random selection of the possible combinations of answers on the individual items of varying difficulty. It is interesting to note that

Walker's concept of "unig" anticipated by some years the notion of what later came to be called a "Guttman scale" (Guttman, 1941); also, Walker's continuum from "hig" to "unig" corresponds to Loevinger's (1947) continuum from maximum "heterogeneity" to maximum "homogeneity." Nevertheless, Walker's concerns did not extend to the question of whether the units of measurement represented by point scores are equal, nor to the question of whether the "unig" condition he described would necessarily reflect true homogeneity of the items with respect to the functions they measured.

A test score obtained as the number of items correctly answered (with possibly an adjustment for "guessing" made by subtracting a fraction of the number of incorrect answers) is called a "raw score." Such raw scores have quite arbitrary meanings. Their direct interpretation would have to depend upon a knowledge of the characteristics of the total assemblage of items—for example, on a knowledge of whether they are generally easy or difficult. Even a conversion of a raw score to a *percent correct* score would help very little. This fact was recognized by early mental testers, with the consequence that elaborate techniques (foreshadowed by Galton, 1885, with his concept of the percentile) were worked out for interpreting scores with respect to the performances of stated groups, for example, ninth graders, 14 year olds, or representative groups of college students. Thus was developed the "normative" interpretation of test scores. The more reputable and responsible publishers of group intelligence tests provided extensive tables of "norms," that is, correspondences between scores and the percentile ranks of those scores in typical groups. Scores could also be interpreted normatively in terms of "standard scores" or "T-scores," that is, values that expressed how far above or below an average a score would be, measured in terms of the standard deviation of the scores of representative groups.

All these normative interpretations of test scores were useful but left something to be desired. A normative interpretation does not really characterize the performance of an individual, for it tells little about the absolute level of that person's abilities. The problem of relating mental test scores to absolute, or quasi-absolute, levels of abilities was never satisfactorily solved in the developmental period of mental testing; in fact, it is a problem even today, although (as will be pointed out) the technology for doing this is now much more advanced.

1.4 *Statistical Methods and Mental Test Theory*

Quantitative and statistical methods have played a large and necessary role in the development of mental tests. It is convenient to organize the discussion of these methods from the perspective of the present day, which would treat them under a number of somewhat separate topics:

1. *Univariate descriptive statistics*: Methods for characterizing the dis-

tribution of the measurements of a given variable over a sample of things (objects, persons, events, etc.) measured.

2. *Inferential statistics*: Statistical methods for making inferences concerning populations from the statistics obtained from samples of those populations.

3. *Bivariate and multivariate descriptive statistics*: Methods for characterizing the joint distributions of two or more measurements over a sample of things measured and for predicting measurements on a given variable from one or more other variables.

4. *"Mental test theory"*: A collection of special statistical methods and theorems that have to do with the reliability and validity of measurements, particularly the kinds of measurements derived from mental tests.

5. *"Factor analysis"*: A collection of special mathematical and statistical methods that provide models for the identification of fundamental dimensions that underlie the observed measurements from mental tests or other sources.

The beginnings of the work in univariate descriptive statistics, and to some extent those in inferential statistics, occurred well before the start of the developmental period of mental ability testing that is now in view here. Most of the major developments in the other topics mentioned above had their beginnings during that period, however. In fact, many of the developments, even in "pure" statistics, were occasioned by the technological requirements of the mental ability testing movement. A detailed account of developments up through about 1928, with emphasis on origins in the 19th century, is to be found in Helen Walker's *Studies in the History of Statistical Method* (1929), actually a revision of her doctoral thesis at Teachers College, Columbia University.

By the time of Galton and Cattell, the theory of the Gaussian law of error—what is now conventionally called the normal probability distribution—had been worked out, and, early in the 19th century, Quetelet, the Astronomer Royal of Belgium, had shown how it could be applied to the analysis of various kinds of vital and social statistics. Concepts of central tendency (indicated, for example, by the arithmetic mean of a series of observations) and of variation (as measured by the "probable error" or the standard deviation) were known. It fell to Galton (1885), however, to make the first wide use of the normal probability distribution in psychological and educational contexts. He made frequency distributions of his mental test results and suggested the use of the normal curve in assigning class grades. He can be regarded as the inventor of what is now called the percentile rank. Univariate statistical methods continued to be developed and applied throughout the period under review here, particularly by Karl Pearson in a series of publications starting in 1894 (Pearson, 1894), G. Udny Yule (1911), Ronald Fisher (1925), and many others, principally in England and often in fairly close contact with

psychologists concerned with the development of mental tests.

Because of their close conceptual relationships with probability distributions, the methods of inferential statistics received much early development in the period under consideration here. In connection with mental tests, they were used mainly in research on the application of tests, for example, in deciding whether there were significant differences between test scores of different sexes, social classes, races, etc. However, even as late as 1935, the "small sample" statistical methods developed by Fisher (1925), "Student" (the pseudonym of one of Fisher's students), and others were practically unknown in America. Fortunately, American psychologists in the early days tended to employ such a conservative standard in testing statistical differences (a "critical ratio" of four times the probable error, corresponding to $p < .007$) that at least it can be said that they only infrequently made "Type I errors"—rejecting the hypothesis of no difference when it was in fact true.

It is in the domain of bivariate and multivariate descriptive statistics that the development of statistical methods had a particularly close association with the development of mental tests, for fundamental to the methodologies surrounding mental tests is the statistical concept of *correlation* or *association*. Reading Helen Walker's (1929, chap. 5) account of its development, one may be surprised to discover that the technical concept of correlation is hardly a hundred years old, having first arisen in the mind of Galton around 1875, in connection with his studies of the inheritance of traits such as the sizes of peas, and yet not appearing in published form until 1886 (Galton, 1886, 1889). (The nontechnical concept of correlation is embedded in ordinary language, as in expressions of the form "the more X, the more (or less) there tends to be Y." The technical concept provides a measure of the degree of relationship between the variables X and Y.) Actually, Galton conceived correlation in terms of *regression*, a term deriving from the fact that the mean heights of sons, for example, tend to "regress" toward the mean height of the fathers. The first exact formula for computing what is today the most prevalent correlation measure, the Pearson product-moment correlation coefficient, was not published by Karl Pearson until 1896, even though, as Walker points out, many earlier mathematical statisticians (e.g., Laplace, Plana, Gauss, Bravais) had been "on the verge" of formulating it. Nevertheless, after the publication of Pearson's work, there was a period of intense activity for about ten years during which most of the basic correlation methodologies were developed, including, for example, the standard deviations and correlations of weighted sums, multiple correlation and regression theory, and theorems on the probable errors of estimates. Practically all this work was accomplished by Pearson and his students and associates, in England.

Strangely, it appears that Galton, the discoverer of the technical concept of correlation, did not recognize its usefulness in the analysis of

mental test results; at least, I have found no indication of his having applied it to the mental test data that he collected at his Anthropometric Laboratory in the years 1884–1890. Instead, Galton applied his correlation and regression concepts to the problem of measuring the role of inheritance in passing traits from generation to generation.

Correlation theory has continued to play an important role in the field of genetics (and in the applications of that field in psychology), but one of the first applications and refinements of correlation theory in connection with mental tests is that of Spearman in his paper in the *American Journal of Psychology*, "The Proof and Measurement of the Association Between Two Things" (Spearman, 1904a). In this and a further paper (Spearman, 1904b), he not only showed how correlation theory could be applied to establish dimensions of mental ability, but also developed several formulas for dealing with the role of error in test results. Spearman saw the correlation coefficient as a measurement of the extent to which two variables may be taken to measure one underlying variable in common. Spearman was the first to apply the correlation coefficient as a measure of the reliability of measurements; these reliability coefficients (though he did not at that time call them that) played a role in his famous formula for "correction for attenuation," a formula for estimating the amount of correlation that would exist between two series of measurements if they could be freed of all error. However, it was not until 1910 that a further derivation of reliability theory was made, independently by Spearman (1910) and William Brown (1910), the so-called Spearman–Brown prophecy formula for predicting the effect of the length of a test (or the number of replicate measurements) on its reliability. (Derivation of this formula is a trivial exercise for present-day students of test theory.)

The work of Spearman and Brown generated what may be regarded as two rather independent subfields in quantitative psychology: one, the specialty known as "factor analysis" and, the other, what has come to be called "mental test theory." Factor analysis is concerned mainly with the identification of the "dimensions" of mental ability, while test theory is concerned with making reliable measurements of mental abilities. (Both factor analysis and test theory can, of course, be applied to other types of measurements.) Actually, there are intimate relations between factor analysis and test theory, but particularly today the specialties tend to be quite separate—they are usually treated in different university courses, taught by different specialists who tend to publish in different sets of journals. And, somewhat surprising to state, factor analysis and test theory are usually treated quite separately from the basic statistical methodology that encompasses the further refinements of correlation theory that ensued from the work of Pearson.

In America, the person most instrumental in disseminating the new discoveries in correlational methods and mental test theory was E. L.

Thorndike. His book, *Introduction to the Theory of Mental and Social Measurements* (1904), contained detailed discussions of statistical methods, including correlation, appropriate for the construction and application of mental tests. Thorndike was even able to take some advantage of Spearman's then recent contributions to reliability theory. The result was a flurry of activity in applying the new methods to a variety of test data in psychology and education. It is rather amusing to note the frequency with which new correlation formulas were "discovered" by American educational statisticians—often these formulas were mere mathematical variants of those already presented by British workers.

By the time of the establishment of the Army testing program in World War I, the principles of basic statistical methods, including correlation, were fairly well understood and played a large role in the selection, refinement, and analysis of the Army tests. In fact, Yerkes's statistical report of the Army testing program (1921) contains one of the early applications of multiple correlation in psychological testing, used to discover the optimal weighting of two or more tests in predicting an independent criterion. Correlations among the separate tests in test batteries were computed in order to see which tests best contributed to the aggregate score and which tests were more independent of the others. By this time, also, simple item-analysis techniques had been developed and were used to arrange items in order of difficulty and to select items that best discriminated individuals on the total score scale.

Probably the person most responsible for the dissemination of psychological measurement statistics in the 1920s was Truman L. Kelley, who for a number of years (1920–1931) was associated with Terman's activities at Stanford University in the further development of the Stanford-Binet scale and a variety of other tests. His book *Statistical Method* (1923) "marked an important milestone in the application of rigorous statistical methodology to problems in psychology, education, and other social science fields" (Flanagan, 1961, p. 343). It summarized nearly all the previous relevant work in statistical method and presented many new contributions from Kelley himself. It was the basic text for a generation of workers in psychological and educational measurement. Another classic work was *The Interpretation of Educational Measurements* (1927), presenting a unified account of mental test theory as it had developed up to that time. Kelley was also a contributor to the development of factor analysis; his monograph *Crossroads in the Mind of Man* (1928) contained some important new derivations.

Statistical methods, test theory, and factor analysis in the 1920s had still not become so highly specialized that they could not be easily encompassed within the expertise of a single individual. One other figure who started to become prominent at this time was L. L. Thurstone (1887–1955). Whereas Kelley had been identified principally with educational

measurement, Thurstone was identified primarily with psychology. Early on, he was a contributor to the field of psychophysics, being responsible for several famous "laws" of comparative judgment (Thurstone, 1927). The author of a theoretical work on the nature of intelligence (1924) and a brief text on statistics (1925a), he was one of the most creative workers in the development of psychological tests. His work in the construction of psychological examinations for college admissions has already been mentioned. All of this work impelled Thurstone's quantitatively oriented mind to seek the elevation of psychological testing to the status of a truly "quantitative, rational" science. Although Thurstone made few fundamental contributions to mental test theory, save perhaps for his so-called absolute scaling methods (Thurstone, 1925b, 1928), he systematically summarized the major parts of test theory for his students in a small monograph, *The Reliability and Validity of Tests* (1931a). This monograph was the inspiration for many later developments in test theory among Thurstone's students and co-workers, of which we will say more in our survey of the "modern" period of mental testing. Thurstone's major contributions were in the field of factor analysis.

In factor analysis, the period from 1904 to about 1931 was heavily dominated by Spearman, who was mainly concerned to establish and validate the two-factor theory he had announced in 1904. From the standpoint of today's theory and technology, Spearman's methodology was quite unsophisticated. It used relatively simple algebraic procedures in order to "factor" a matrix of intercorrelations. Because of the unavailability of efficient computing procedures, Spearman and his students generally used small sets of variables—rarely more than 20 or 25, and more commonly 10 or 12. Progress was relatively slow. Techniques for establishing new group factors were primitive. A major part of Spearman's scientific output was of a polemical nature, arguing against critics of his rather narrow two-factor theory. It may be said that scientific progress in the analysis of mental abilities was held back by the lack of adequately refined mathematical and statistical methods for such an analysis. Nevertheless, Spearman was sufficiently optimistic about the prospects of new findings to establish, in the early 1930s, with Karl Holzinger at the University of Chicago, a large "Spearman-Holzinger Unitary Trait Study" to search for "group factors" of mental ability.

By this time, Thurstone was literally just down the street from Professor Holzinger at the University of Chicago, yet the two never collaborated directly. Thurstone's major contributions came through his recognition that the factor analysis of correlation matrices required sophisticated methods of determinants and matrix algebra, and although the basic theorems had long been known to mathematicians (e.g., Bôcher, 1907), they were known to few psychologists. Thurstone set about learning them, under tutelage, and by 1931 he was able to sketch a new model for factor

analysis that was, in effect, a generalization of Spearman's model to multiple factors (Thurstone, 1931b). A more complete version of this model was published several years later, *Vectors of the Mind* (Thurstone, 1935), and, within a year or so, Thurstone, his wife Thelma, and their students were starting to assemble a major data base to test out these new formulations. The results of this study were published as a monograph, *Primary Mental Abilities* (Thurstone, 1938), setting the stage for an entirely new phase in the development of theories of mental abilities, discussion of which we reserve for our treatment of the modern period of mental ability test development.

Advances in computational technology have had a large role in the development of mental ability test methodology. The computational methods available to people like Galton and Cattell were primitive in the extreme; aside from the slide rule, there were practically no aids to hand computation. The period after Thorndike's introduction of statistical methods to American psychologists saw the production of numerous "worksheets" for computing correlation coefficients—many of them exceedingly clever—that facilitated computations but that still relied essentially on hand tabulations and mental arithmetic (Otis, 1923; Toops, 1921). Tests had to be scored by hand, although fairly efficient methods utilizing "scoring keys" were devised early in the century. Mechanical aids such as machines for basic arithmetic operations were introduced only gradually, and their use was cumbersome and error prone. Hollerith machines using punched cards had been introduced by the U.S. Census Bureau early in the century and received some use in the Army examining program of World War I, but their use was limited to providing data for simple hand-tabulations. Even as late as 1935, their uses in performing any elaborate statistical calculations were extremely limited, although Warren and Mendenhall (1929) had developed a method for computing correlations partly by tabulating machine. From the late 1920s until well into the 1950s, the mechanical desk calculator was the essential tool of the statistician. It is a practical certainty that many published research results contained egregious errors that were the fault of the investigators, not the printers. Of course, errors occur even with present-day electronic facilities, but these errors are rarely of a primitive computational kind. The more basic point that should be made here is that the lack of efficient methods for collecting, tabulating, and analyzing statistical data seriously impeded progress during the developmental phases of the mental ability test movement.

1.5 Test Construction, Standardization, and Validation Methods

Previous sections have outlined the theoretical and methodological developments that occurred in the early phase of the mental ability test movement. To examine the actual impact of these developments on the

construction and use of tests, it may be helpful to take a look at several group mental ability tests that were in wide use at the end of the period. We shall examine a form of the *Henmon-Nelson Tests of Mental Ability* that was designed for use with elementary-school children, a form of the *California Tests of Mental Maturity* that was designed for use with secondary-school students, and one of the annual editions of the *American Council on Education Psychological Examination*, designed for use with college students. All three tests were constructed and published around 1935.

The *Henmon-Nelson Test of Mental Ability, Elementary School Examination* (Henmon & Nelson, 1932-1935) is a group test designed for students in elementary school, Grades 3 through 8. It was first published in the years 1932 through 1935 and early achieved wide popularity because of the simplicity of its format and scoring (from the standpoint of the test user), the small amount of time required to administer it (30 minutes), and its low cost. Actually, this test is one of a set of three, each test in the set being designed for a different range of grade levels. (The other two levels are a *High School Examination* for Grades 7 through 12, and an *Intelligence Test for College Students*, by the same authors.) The test has been revised several times; one revision was published in 1957 by Tom A. Lamke and M. J. Nelson and a further revision was published by Joseph L. French in 1973.

The test itself consists of a series of 90 items printed in a booklet. At the outset of the test, the examinee is given examples and practice exercises for the several types of items he or she will encounter in the test, such as vocabulary, sentence completion, "classification" (finding which one of a set of stimuli cannot logically be grouped with the others), series completion, analogies, and arithmetic problems. All items in the test proper are in multiple-choice format; the student has to find the one correct or "best" answer from the five choices offered, indicating the choice by marking in a particular space. The marks are registered through carbon onto a grid underneath; the test can then be scored relatively easily (though manually) by counting the number of marks that fall in designated positions on the grid. (In this way, the test makes use of the Clapp-Young Self-Marking Device, patented in 1929.) The items are arranged in order of "difficulty"; at the same time, the several types of items appear in more or less random order, it being assumed that the student has been adequately instructed at the outset of the test in how to do each type of item.

Many of the item types require the student to be able to read printed English; others employ numerical or geometrical ("spatial") material. After the introduction to the test, the student is given 30 minutes to do as many items as he can—90 items in 30 minutes, or 3 items per minute on the average. It is a rare student who does all the items within this time

limit. The raw score is the number of items answered correctly within the time limit.

There are three forms of the test, and according to the test manual they may be regarded as equivalent—i.e., as yielding scores that could be expected to be approximately the same regardless of which form is administered. Tables are supplied whereby the test user can easily convert the raw scores to mental ages and thence to IQs. Normative data, based on a total of about 5,000 cases sampled from various elementary schools around the country, are given for each grade for which the test is designed. Separate reliability coefficients, found by the odd-even technique and the Spearman-Brown formula, are reported for each age and each grade. These reliabilities are all high, being consistently in the upper .80s or in the .90s. It should be remarked, however, that such reliabilities may well be inflated by the fact that there is a speed element in the test. When a test is "speeded," there will be variation in the numbers of items attempted by the examinees, and when scores on the odd items are correlated with the even items, a consequence of such variation in numbers of items attempted is that a rather high degree of correlation is inevitable. The use of odd-even reliability coefficients may therefore give a falsely favorable impression of the actual accuracy of the test scores. The authors failed to recognize this fact; they also failed to report equivalent-form reliabilities, something which they could easily have done.

The "validity" of the test as a measure of intelligence or mental ability was ascertained by correlating scores of the test with other measures presumed to reflect intelligence. Correlations of the scores with results of other widely known group intelligence tests are reported as ranging from .54 to .92. Such values are regarded as being generally satisfactory, particularly the upper range of these values (9 of the 19 correlations reported were above .80). The authors also imply that the scores should correlate well with scholastic performance, since they report that in the process of selecting items for the test, only those items were retained that "proved to differentiate between pupils of known superior and known inferior mental ability." They fail to say exactly how pupils' superior or inferior abilities were "known," but presumably the criterion was scholastic success as reflected in teachers' grades.

In their manual (Henmon & Nelson, 1932-1935), the authors make various suggestions concerning how the test results can be used. They state their belief that the test measures "those aspects of mental ability which are important for success in academic work." The scores are to be interpreted as measures of general mental ability but not as indices of school achievement specifically taught in a common curriculum:

Although knowledge is required to answer the items, they were selected to reflect ability to profit from exposure to learning in a variety of situations. High performance on the test requires the efficient utilization of verbal and

numerical symbols and the ability to acquire and retain information in common symbol form for use at later times in the solution of verbal, quantitative and abstract reasoning problems. (Manual)

They believe that the results will be most valuable to teachers when the pupil's performance on the test is compared to his performance in school work; in this way the teacher can arrive at a judgment as to whether the student is working up to capacity. The implication is that a student with low scores on the test *and* with low scholastic performance is probably working up to his rather limited capacity. A student with high scores on the test and *low* academic achievement is one that would need help and special attention. If a student makes low scores on the test, but still does quite well academically, it is recommended that the scores on the test be "checked" through the administration of individually administered tests such as the Stanford-Binet.

Several experts in psychological measurement commented upon the test in a kind of "consumer's report" compiled by Buros (1941). Anne Anastasi, who later became one of the nation's leading authorities on individual differences (Anastasi, 1958), remarked that "the use of a single score based on a hodgepodge of different types of items could be questioned, if one wanted a particularly fine or discriminative measure. But like other tests of its kind, this scale serves the practical purposes of (a) preliminary rapid exploration, and (b) rough classification of broad groups" (Buros, 1941, p. 221). Howard Easley stated: "The content and standardization of these three tests seem as satisfactory as those of the better group tests of intelligence, but not strikingly more so" (Buros, 1941, p. 222). J. P. Guilford, later to become the originator of a widely known theory of intelligence (Guilford, 1967), proclaimed: "An examination of these group tests and the manuals and scoring keys that accompany them impresses one immediately with the kind of care and expertness with which one wishes all tests were constructed" (Buros, 1941, p. 222). Although they had certain reservations about details, all these experts were willing to accept the test as a well-constructed, highly useful measure of mental ability.

The test can be appraised as the natural resultant of trends that had been initiated by Binet and other early developers of mental tests. The kinds of items included in the test can be regarded as group-test derivatives of the sorts of tasks included in the Binet scale—simple problems involving the understanding of words and verbal expression, the manipulation of numerical and spatial information, and the recognition of logical distinctions and classifications. It could be argued that the materials in the test items were not specifically taught in the school curriculum, but that successful school learning would require the student's ability to perform problems with such materials. (A few items obviously require certain kinds of specific information, but this information is of such a general and

widely disseminated character that it would be only the less "able" students who would be less likely to have acquired it.)

The construction and standardization of the *Henmon-Nelson Test of Mental Ability* obviously involved various technologies that had been fashioned throughout the early days of the testing movement. The items had been selected from pools of items that had been submitted to teachers, tried out on a number of occasions, and evaluated for their overall "difficulty" (as reflected in the proportions of students in different grades able to pass them) and for their power of "discriminating" among groups of students of "known" mental ability. Reliability and validity coefficients had been duly computed according to Pearson and Spearman's formulas, and norms had been compiled on the basis of fairly representative samples. The format was carefully devised in the light of experience with printed tests of this type, and the scoring and administration of the test had been honed and refined to a point of high efficiency—as much efficiency, that is, as could be achieved prior to the development of test-scoring machines.

Another example of a group intelligence test developed during the early period is the *California Test of Mental Maturity, Advanced Series*, 1937 edition (Sullivan, Clark, & Tiegs, 1936–1939). This test was designed for pupils from Grades 7 through 14. According to the authors' manual, its "primary purpose is to make for each pupil a diagnostic evaluation of those mental abilities which are related to, or determine, his success in various types of school activity in order that the teacher may utilize this information directly in aiding students who have learning difficulties." In contrast to the test we have just considered, it yields not a single score but a whole series of scores that can be plotted on a "diagnostic profile." It is

based on the philosophy, researches and important inferences of outstanding leaders in this field, as well as the work of the authors. The analytical comparison of the various sections of the test indicates a definite central factor, yet the same analysis reveals a specificity for each test sufficient to justify its inclusion as a measure of a more or less unique factor. . . . This series of tests is unique in that each battery is preceded by tests of visual acuity, auditory acuity, and motor coordination, the purpose of which is to detect those pupils with defects sufficiently serious to prevent obtaining a valid diagnosis of mental maturity with the remainder of the test. In general, the mental maturity test proper samples memory (immediate and delayed); maturity of apperceptive processes; spatial relationships; and logical and mathematical aspects of reasoning. Certain tests are presented in both verbal and non-language form in order to obtain a separate evaluation of each. . . . The authors of these tests believe that the multiple-factor [sic] theory of intelligence comes nearer to explaining observable phenomena than does the strong central-factor theory alone . . . [and] believe that progress in determining the nature of mentality and the value of tests of mental maturity is dependent largely upon further studies in factor analysis which employ analytical and statistical techniques. (Manual)

The battery consists of 16 tests and takes a total of one and a half hours to administer. (Generally, it would have been administered in a series of class sessions on different days.) The first three tests, presumably measuring visual acuity, auditory acuity, and motor coordination, are not to be counted in the total scores, but their results are to be examined by the teacher for insight into whether the remainder of the test is to be regarded as valid for the individual pupil. The 13 remaining tests are generally quite short but exhibit considerable ingenuity in their construction. Only some of them utilize a multiple-choice format. The scoring is to be done by hand, using a series of answer keys. (Later editions of the battery were converted to machine-scoring formats.) In general, the subtests represent extensions of Stanford-Binet tasks to a group testing situation. For example, Test 4, of "immediate memory," represents an ingenious procedure for testing digit memory span in a classroom test. Test 5 involves delayed recall of paragraph meaning: A story is read near the beginning of the test and memory for its contents is tested at the end of the test. Three tests of "spatial relationships" require the student to "sense right and left" (determining whether pictured hands and similar items are "right" or "left"), to detect whether pairs of geometric figures are rotated or "flipped over," or to find the shortest path through a maze (thus exhibiting "foresight in spatial situations"). Various "reasoning" items are presented mainly in pictured form; for example, in a test of "opposites," the student has to find the "opposite" of a picture of a cake (it turns out to be a picture of a lemon, because while a cake is sweet, a lemon is sour). Even analogies are presented in pictorial form: A *speedometer* is to a *car* as a *thermometer* is to a (*rose, sun, turnip, airplane*), where all the italicized words are pictured. Other reasoning items include number series and various types of numerical problems (e.g., how to make up \$5.51 with exactly 15 coins). The last three tests are tests deliberately using verbal presentation, arithmetic problems, syllogistic inferences, and vocabulary.

The test manual gives a variety of data on the statistical characteristics of the results, chiefly split-half reliabilities and tables of norms. The authors do not seem to be aware that the split-half reliabilities given for each grade, all quite high in magnitude, are probably inflated because of the speededness of the tests. (The authors insist, however, that "this is a power rather than a speed test.") Intercorrelations among the tests are not reported. It would have been desirable for the authors to do so in order to enable the user to assess the extent to which the separate tests measure distinct abilities.

The manual presents much advice to teachers on how to interpret results. For example,

A student who reveals high ability in Tests 9-13 inclusive, and who rates low in Tests 14-15, may lack reading ability. . . . Where disability is not due to some remediable cause, a student who cannot detect differences and similarities or sense relationships is simply missing important elements of his environment; if in addition he fails to sense quantitative relationships and fails in manipulating both quantitative and non-mathematical ideas, he is destined to live on a low plane intellectually. Such students tend to remain in a simple concrete world; they do not generalize easily, must often be given their principles and generalizations as facts, and must depend upon memory rather than ability to deal with ideas. Notwithstanding these disabilities, some students do well in situations involving common sense (because of experience) while more brilliant inexperienced students will fail. Students who do poorly on these tests cannot go far beyond limits set in general by their mental ages. While this is true of the types of activities which constitute the ordinary school program, many students who do poorly on these tests nevertheless have talent in music, art, or some other special field. It is well to follow this general test with tests of specific aptitude.

Maturity of the apperceptive processes with reference to ideas and growth of meaning is revealed by Test 16 [of Vocabulary]. The verbal factor is recognized as an important and more or less independent element of mental maturity. Meaning is added gradually and skill in using verbal symbols is primarily a matter of training. . . . A high rating in this test is evidence of the capacity to profit by experience, resulting in a high degree of development of the verbal factor. However, a low or even an average rating is not a definite proof of native limitation of the verbal factor; favoring circumstances and specific training in language may have been lacking.

After the total test scores have been recorded and the relative rating in language and non-language factors obtained, the teacher will have revealed to her whether or not the student will probably learn best through language or non-language activities. (Sullivan, Clark, & Tiegs, 1936-1939, Manual)

The important influence of Binet, Terman, and others on the development of this test is evident. In fact, its chief author, Elizabeth Sullivan, had worked with Terman and published her research on applications of the Stanford-Binet scale (Sullivan, 1926). The test already shows, however, evidence of the then new indications that mental ability is not to be described in terms of a single "general factor" but in terms of a series of differentiated traits. Just how much of the technology of mental testing is incorporated into the test is unclear; there is no indication in the manual about the use of item analysis or other test-refinement procedures.

Nevertheless, one commentator (Raymond B. Cattell) in Buros's (1941) compilation was able to say:

These tests are exceedingly well designed from the point of view of adaptation to school needs and the convenience of the teacher. All the data regarding consistencies, standardization, correlation with school progress, etc., that one could reasonably demand, are clearly presented in the handbook of instructions. . . . An admirable feature of this test is the courageous manner in which the authors come out into the open regarding the purpose, principles, and theory of test design. (p. 208)

A little later, he comments:

They offer a profile which "analyzes and summarizes the various factors which are measured by the test situations," and claim that this "reduces the mystery which has surrounded the meaning of mental age and intelligence quotient." This attempt to produce for special consumption a "psychology without mystery" ends by appearing to the psychologist to be "mystery without psychology." No proof is offered that these subtests do, in fact, test independent factors or that one is justified in generalizing from them to performances in everyday life which happen to have the same verbal label. (p. 208)

Another commentator (F. Kuhlmann) is more critical:

We do not believe there is much merit in labeling tests as regards functions measured, as the authors have done; first, because it cannot be done correctly by inspection; and second, because these labels are not of much value until we know also how these functions enter into school achievement in different school subjects. Also, when a battery is divided into several different measures the tests assigned to measure any particular function tend to become inadequate in number and range to do so reliably. It would be hazardous, indeed, to conclude from the score on two brief tests that a child has a poor memory, for example. It seems to be implied also that the child mind is simply the adult mind in miniature, so that all tests should measure the same function at all ages. . . . The authors' distinction between language factor and nonlanguage factor tests is also somewhat misleading. Language enters both, the real distinction being that in the former the child has to read test material, while in the latter he is told what to do with picture material and, with a few exceptions, no reading is involved. (Buros, 1941, p. 209)

What Kuhlmann is complaining about is a problem that has persisted up to the present day: While total scores on mental ability tests can be shown to be rather highly correlated with total progress in school, it is very difficult to prove detailed relationships between separate kinds of mental ability and particular activities in school learning. He also points to a difficulty that has cropped up in all "multifactor" batteries of mental ability: It is hard to obtain reliable measures of different kinds of mental ability within very short time limits.

One more widely used group intelligence test deserves to be examined, this one issued by the prestigious American Council on Education. A "psychological examination" constructed by L. L. Thurstone and Thelma Gwinn Thurstone was issued annually by ACE from 1924 to 1954; for present purposes we select the one published for 1935, the 12th edition of the test. According to Thurstone and Thurstone (1936), it was ordered by 493 institutions, testing a total of 189,506 students. The article by the Thurstones reports data on the distributions of scores for 58,402 students from 266 colleges, but, unlike the case in former years, "in order to discourage the use of test results as the basis for comparative ranking of institutions, this year the names of the colleges have been deleted."

The test was issued in both hand-scored and machine-scored editions;

while the test contents of these editions were slightly different, the authors claim that scores on them had been shown to be equivalent. The hand-scored edition, at any rate, had five parts:

1. *Completion* (40 items, 10 minutes). The examinee has to write answers to items such as "Burnt sugar, used for coloring and flavoring, is called (7 letters)." [Answer: caramel]

2. *Arithmetic* (20 items, 20 minutes). The examinee has to supply the answers to arithmetic reasoning problems such as "A freight train traveling at the rate of 15 miles per hour requires 80 minutes between two stations. What time does an express train going 48 miles per hour require for the trip: — hrs. — min. — sec."

3. *Artificial Language* (30 items, 13 minutes). After vocabulary, grammar rules, and examples for an artificial language are given, the examinee has to evaluate the correctness of translations of sentences like "Slupup girdigro paresba."

4. *Analogies* (29 items, 10 minutes). In each item, the examinee has to detect a relationship exemplified in two geometric figures and find which of several other figures exemplifies this same relationship to a third figure.

5. *Opposites* (33 items, 7 minutes). Each item presents four words; the examinee has to find which two are either the same or opposite in meaning; for example, in the group (many, ill, few, down) *many* and *few* are opposite in meaning. Many of the words used in the test are relatively infrequent and unfamiliar.

No data are presented concerning the reliability of this instrument, nor on the manner in which it was constructed. Buros's (1941) compilation contains, however, a long series of citations of research articles by various authors reporting on correlations between the various annual editions of this test and scholastic success in college. It would appear that many of these correlations were sufficiently high to justify continued use of these examinations for placement and other purposes in colleges. Only in a later article by the Thurstones do we find validity coefficients reported. For data collected in 1933 and 1934, correlations of the ACE Psychological Examination Score with average scores in introductory courses in various subjects were about .50. Furthermore, the Thurstones had started to recommend the computation of separate L ("Linguistic") and Q ("Quantitative") scores. The L score was to be computed by adding scores on the Completion, Opposites, and Artificial Language subtests, while the Q score was to come from the Arithmetic and Analogies subtests. According to the Thurstones, "the two constellations are correlated (.55) but do not represent primary factors. Both of these composites are complex as to their primary factorial composition" (Thurstone, Thurstone, & Adkins, 1939).

Our examination of these three tests bears out the statement made

above that during the early developmental period, and by about 1935, the science of mental testing progressed from its crudest beginnings to a stage of high development. The procedures developed by Binet for gauging mental growth by individually administered tests had been adapted for use in group intelligence tests. Theories of intelligence had been formulated, and technologies for constructing, standardizing, and validating mental tests had been developed, utilizing advances in statistical methods that occurred during this period. There was widespread acceptance of the notion that mental ability as measured by intelligence tests played an important role in students' school success. By 1935 there were the beginnings of interest in the differentiation of mental abilities, but there were as yet few indications of the role of such differential mental abilities in school performances.

1.6 Practices in the Use of Tests

As we have seen, by 1935 intelligence tests of various sorts had been widely applied in a variety of settings. These uses were reviewed in a number of textbooks designed for psychologists, teachers, and other practitioners. One example is a work by Pintner (1931), a professor of education at Teachers College, Columbia University, and an author of intelligence tests. Pintner points out that in the beginning the development of psychological tests was "closely bound up with the study of mental deficiency and abnormality" and shortly afterward with the study of juvenile delinquency and of the dependent child. "The next striking phenomenon," he says, "was the appearance of the very bright or superior child. . . . It is doubtful whether the very decided superiority of some children was really appreciated before the advent of the psychological examination" (p. 226). With the coming of the group test method,

it became possible to test large numbers of school children. This brought in the period of school surveys on a large scale with their direct and important influence for the purpose of instruction. At the same time the group method was extended to other groups of individuals, soldiers in the army, prisoners, college students, and the like. . . . We note also the extension of the use of tests to special groups of individuals . . . the blind and the deaf.

Later he continues:

During all this development the use of intelligence tests in the study of racial differences has been going on. The most thoroughgoing studies have been made on the negro in America. . . . The possibilities in this line of work are very great, although there are certain inherent difficulties that have not yet been fully overcome. . . . The field of industry and commerce is one of the latest fields in which intelligence tests have been adopted. (p. 227)

Devoting separate chapters to uses of tests in different situations, Pintner notes that, in the elementary school,

the chief practical uses of tests up to the present time have centered around their value for the purpose of classifying children into more or less homogeneous intelligence groups, and also for predicting their future success in school work. These two purposes are intimately bound up with each other. Classification in homogeneous groups is justifiable because intelligence correlates highly with school success, and therefore, the more homogeneous the group the more likely are the children in the group to advance together at about the same rate, be that rate relatively fast, normal, or slow. (p. 239)

McClure (1930) had in fact found, through a questionnaire sent to a number of large city school-systems, that the most important use of psychological tests was for sectioning classes into homogeneous groups. Nevertheless, evidence that homogeneous grouping of children with respect to intelligence actually made for more effective educational programs was very scarce. Most of the reports, according to a review by Rankin (1931), were negative. The failure of homogeneous grouping, Rankin thought, might be attributed to the probability that this practice was not ordinarily accompanied by appropriate modifications of curriculum materials and instructional practices for different classifications of learners. Keliher (1931) had attacked the whole practice of homogeneous grouping on the ground that it did not square with an educational theory that conceives education as growth and takes the whole child into account. Further, Keliher pointed out, as did many others, that children are so specific in their abilities that it is difficult if not impossible to achieve true homogeneity of grouping. Despite all this, it appears that homogeneous grouping was widely practiced by 1930. (In a later section of this chapter, we shall see that the practice continued for many years.) Pintner (1931) confirms this in his summary statement:

It is evident from the survey in this chapter of the use of intelligence tests with school children that such tests are at present being used very widely and for many different purposes. They have not come to be considered an integral part of every school system, but the time is not far distant when they will be considered as essential for the health, happiness and advance of every child.

At present tests are being used for the purpose of classifying children into more or less homogeneous groups. . . . Most reports of such classifications are favorable, and justice is being done to all types of children by this procedure. . . . A natural corollary to this movement of sectioning according to ability will be the gradual differentiation of the course of study in terms of capacities of the various groups of pupils. Little has been done in this respect, but the indications are that attempts in this direction will be made now that the individual differences have been so clearly demonstrated. (pp. 268-269)

At the high-school level, Pintner was able, at the time he was writing, to point to the fact that selection for admission to the high school "was, and still is, partly on the basis of intelligence," even though the percentage of the population of high-school age enrolled in school had increased from about 12% in 1910 to about 47% in 1926. He seemed to express concern in his statement that "as the proportion of children going to high school

increases the average intelligence decreases" (p. 275). Mort and Featherstone (1932), in a review of the uses of psychological tests in schools, noted that use of such tests for admission to high school "doubtless underlies much of the investigation that has been carried on" with respect to the predictive value of intelligence test scores for success in high school, but they remark that as compared to the question of admission requirements for colleges,

there is less significance in the question of initial estimate before admission [to high school], for it is coming to be generally conceded that the high school should be available for all who choose to come. Therefore, at the high-school level, psychological tests are of value, for the most part, in guidance after admission. (p. 302)

Pintner (1931) was nevertheless able to point to many studies showing substantial correlations between intelligence test scores and high-school success, both in particular courses and as reflected in overall averages. He also noted that the average intelligence test scores of students in different types of curricula varied, students in "classical" or "academic" courses having higher average scores than those in "general," "vocational," or "commercial" curricula. "The general picture resulting from these various studies is that of a decrease in intelligence as the course becomes less abstract or more practical in nature" (p. 284).

At the college level, the use of intelligence tests for admissions purposes appears to have begun shortly after World War I. Pintner states, "During the war, and for several years thereafter, Army Alpha was used widely in colleges and universities" (p. 295). He cites a 1923 report by Wood (Pintner, 1931), who stated:

In 1919 the Office of Admissions at Columbia University inaugurated an experiment for the purpose of discovering the value of the Thorndike College Entrance Intelligence Examination for High School Graduates as a criterion for admission to Columbia College. From an attitude of healthy if not severe skepticism toward the use of intelligence tests for this purpose, the whole college administration came, within the space of two years, to consider the intelligence tests as an indispensable part, not only of the admission machinery, but also of the administration of the college in the Dean's Office. (p. 295)

Presumably, the favorable attitudes toward intelligence as a selection criterion resulted from findings concerning the correlations between mental ability scores and scholarship records. Wood reported, for example, a correlation of .67 for 111 students who remained in college for two years. (This correlation is, it should be noted, somewhat higher than those usually obtained for students already selected partly on the basis of scholastic aptitude tests; it appears that Wood's group was largely an *unselected* group.) Summarizing coefficients of correlation found between intelligence scores and college academic records, Toops (1926) reported a median coefficient of .46 based on reports from 43 colleges; Pintner

reported later data with a similar magnitude of coefficients. It will be recalled that Thurstone, Thurstone, and Adkins (1939) reported even higher coefficients. In any case, from 1919 to 1935 there was increasing use of intelligence or "scholastic aptitude" tests in connection with college admission, based on the finding that there were very substantial relationships between scores on these tests and performance in various types of academic college courses. Since records of scholastic aptitude scores were maintained chiefly only in college admissions offices, it is probable that there was little "contamination of the criterion" involved in these findings; that is, it is improbable that professors giving academic grades were informed of the aptitude test scores, and thus the correlations are not spuriously biased towards the high side.

Apparently, intelligence test scores were rarely used in colleges and universities for purposes other than admissions. There are few reports of their further use in sectioning, failing, or counseling students. Nevertheless, some interesting correlates of intelligence were found: Younger students were found to have higher intelligence test scores (a finding that can also be observed in elementary and secondary schools, particularly when students are promoted partly on the basis of performance); also, the more intelligent students reported spending somewhat less time in study (Crawford, 1929).

Throughout the school system, from elementary school to the university, the use of intelligence tests seems to have been predicated on the assumption that their scores reflected mainly innate or at least relatively unalterable characteristics of students having to do with their capacity to do school work. Although it was noted that average scores were correlated with demographic variables such as socioeconomic class, race, urban/rural environment, etc., there does not seem to have been any serious consideration of whether children's home backgrounds, or even their schooling, would have had any important influence on their performances in mental tests. Thorndike et al. (1927) had, to be sure, recognized the difficulty of distinguishing between types of tests that measure mental "maturity" and those that measure the effects of training. But the question of whether test scores were "biased" by cultural factors, for example, was hardly ever raised during the developmental period of the mental testing movement. The general attitude among educators and the public toward intelligence tests during this period seems to be well reflected in the following quotation from Pintner's (1931) text on intelligence testing.

Intelligence tests have done much to show that all children are not created free and equal with respect to their mental abilities. A child's abilities are determined by his ancestors, and all that environment can do is give opportunities for the development of his potentialities. It cannot create new powers or additional abilities. This, then, is the main function of education, to measure the inherited capacities of the child and to so arrange the environment as to

give full opportunities for all these capacities to develop to the uttermost. (p. 520-521)

An overall appraisal of the developmental period of the mental testing movement would have to observe that while major technological advances had occurred to produce a new type of social instrument that in the climate of the times received wide acceptance, the exact nature of this instrument, and its true significance, was largely unexplored. It is evident that by 1935 educators and psychologists were rather well satisfied with what had been created, but at the same time they were highly optimistic about the possibilities of its future development.

2. THE MODERN PERIOD

2.1 Overview

"The pioneering stage in psychometrics may be considered to have ended in the 1930s," says Philip DuBois (1970, p. 112) in his history of psychological testing. Or, as Walter Monroe (1945) put it, it was in this period that measurement moved from "adolescence to maturity." The beginning of the modern period, as we are calling it, may with some justification be placed in the year 1935, the date of the founding of the Psychometric Society by a group of scholars who had been called together by Louis Leon Thurstone and some of his colleagues at the University of Chicago (for details, see Dunlap, 1942). It was Thurstone and the leaders of the Psychometric Society, notably Paul Horst, that were primarily responsible for the origination and dissemination of theoretical and technological developments that had a major impact on mental testing during this period.

We have seen that by 1935, group mental testing had become widely adopted throughout all levels of education. By this time, also, it was beginning to be utilized in business and industry. (It had, of course, been used extensively in the military during World War I, but by 1935 mental testing in the military had become dormant.) Mental ability testing had become an established subject matter for research and teaching in nearly all the major universities, both in departments of psychology and in departments of education. Among the university centers of research in mental ability testing that were particularly prominent at the time were those at Columbia (especially at Teachers College), Stanford, Harvard, Minnesota, Ohio State, Iowa, California, and Chicago. (Centers of research in England and Scotland were also still very active.) Thousands, or more like millions, of "intelligence tests" were administered every year in elementary and secondary schools, often with a kind of blind faith in the scores. Increasingly, colleges and universities were using mental ability test results in their admissions and placement processes, although it was

not until 1937 that the College Entrance Examination Board established scholastic aptitude tests as a regular part of its admissions testing program.

Most of the mental testing that was being done in 1935 was based on the view that "intelligence" is a more or less unitary trait. Most tests of intelligence yielded a single score for an individual—expressed usually in terms of an "intelligence quotient" or some derivative of the intelligence quotient. The common view—even at most of the university research centers named above—was that intelligence test scores reflected some largely innate capacity of the individual to think in abstract terms, to learn in school, and to adapt to the requirements of an increasingly complex technological society. Intelligence tests were valued as scientific measurements to the extent that they showed high degrees of reliability and validity as indicated by the techniques developed by Spearman, Kelley, and other early scholars in mental testing.

Presaged in the later years of the testing movement's early period, but taking full form only in the 1940s and 1950s, two striking technological developments were initiated around 1935, chiefly by Thurstone and his colleagues at the University of Chicago: (a) multiple-factor analysis, a mathematical technique whereby it was clearly shown that "intelligence" is not a simple unitary trait; and (b) the refinement of the statistical theory of mental tests, leading to new ways of constructing mental tests and studying their measurement characteristics. Both of these developments had a substantial impact upon the development and construction of mental tests during the modern period. We shall be looking more closely at these developments in later sections, but for the present we wish to give a general account of mental testing activities during this period.

The history of testing throughout the modern period can be traced through the publications of a remarkable husband-and-wife team based at Rutgers University, in New Jersey. It was shortly before 1935 that Oscar K. Buros, then an assistant professor who had been trained at Teachers College, became disturbed with the laxity of the standards by which mental tests were being published and distributed. He decided to initiate a series of publications that would constitute a kind of "consumer's report" on mental and educational tests. The first two of these compilations were published by the School of Education at Rutgers, but by 1941 Buros and his wife Luella were able to establish their own independent nonprofit publishing house (which they later called the Gryphon Press). Increasingly weighty tomes were issued in 1941, 1949, 1953, 1959, and 1972, with reviews of literally thousands of mental and educational tests; even at this writing, a final compilation is in preparation, the *Eighth Mental Measurements Yearbook*.

The objectives of the Buros series, as described in the introduction to the *Seventh Yearbook* (Buros, 1972), have been:

(a) to provide information about tests published as separates throughout the English-speaking world; (b) to present frankly critical reviews written by testing and subject specialists representing various viewpoints; (c) to provide extensive bibliographies of verified references on the construction, use, and validity of specific tests; (d) to make readily available the critical portions of test reviews appearing in professional journals; and (e) to present fairly exhaustive listings of new and revised books on testing, along with evaluative excerpts from representative reviews which these books receive in professional journals.

But the Buroses attached much importance to other objectives of a "crusading nature":

(f) to impel test authors and publishers to publish better tests and to provide test users with detailed information on the validity and limitations of their tests; (g) to inculcate test users with a keener awareness of the values and limitations of standardized tests; (h) to stimulate contributing reviewers to think through more carefully their own beliefs and values relevant to testing; (i) to suggest to test users better methods of appraising tests in light of their own particular needs; and (j) to impress test users with the need to suspect all tests unaccompanied by detailed data on their construction, validity, uses, and limitations—even when products of distinguished authors and reputable publishers.

Nevertheless, they say:

Our success in attaining the last five missionary objectives has been disappointingly modest. Test publishers continue to market tests which do not begin to meet the standards of the rank and file of MMY and journal reviewers. At least half of the tests currently on the market should never have been published. Exaggerated, false, or unsubstantial claims are the rule rather than the exception. Test users are becoming more discriminating, but not nearly fast enough. (pp. xxvii-xxviii)

These quotations suggest that while the modern period saw the introduction of many innovations and refinements in the theory and practice of mental testing, these innovations and refinements were not utilized and applied as widely and as uniformly as they might have been. In fact, many tests published and circulated during the modern period did not even meet the criteria of excellence that had been established by 1935. The Buroses may be given major credit for helping the field to police its own activities, but even the efforts of special committees of professional organizations such as the American Psychological Association (APA) and the American Educational Research Association (AERA) in publishing quasi-official technical standards for tests (APA, 1954, 1966) were not sufficient to ensure that all test-development activities conformed to these standards. Still less did the run-of-the-mill intelligence tests of the period take advantage of the technological refinements introduced during the period.

We can, however, discern the influence of Thurstone and other leaders in psychometrics in many published tests and in a variety of research and testing programs that appeared or occurred during the modern period.

The prototype of the many "multifactor batteries" that were developed during the period was the Thurstones' own *Primary Mental Abilities* (PMA) battery, an experimental edition of which was published as early as 1938 (Thurstone, 1938b). This battery was based on the findings of Thurstone's first major study of mental tests employing the multiple-factor analysis method (Thurstone, 1938a), a study in which at least seven somewhat independent "factors" of mental ability had been identified. The PMA battery offered tests of seven factors, designated by capital letters and described as follows: P, perceptual ability; N, numerical ability; V, verbal ability; S, spatial-visualizing ability; M, memory; I, induction or generalizing ability; and D, deductive or reasoning ability. More formal versions of the PMA Battery were published subsequently (Thurstone & Thurstone, 1946-1958), for several age levels, but it may be remarked that these tests were never subjected to as much careful refinement, standardization, and validation as they might have been. Reviewers of these tests in the *Buros Yearbooks*, at least, expressed considerable dissatisfaction with the measurement characteristics of these tests. The importance of the PMA battery was that it first brought into general awareness the concept of differentiable mental ability factors.

Mental testing took new strides in the personnel programs of World War II. As had been the case in World War I, many well-known psychologists contributed their talents and energies to the speedy development and validation of tests suitable for use in the selection and training of millions of armed forces personnel for their varied tasks. At the foundation of the testing programs of the several armed forces components were one or more "general classification" tests that resembled the group intelligence tests of earlier periods. For example, in the Army there were developed the Army General Classification Test (AGCT) and the Officer Qualifying Test (OQT); in the Army Air Force, an Army Air Force Qualifying Test. These tests were fashioned along familiar lines, and generally yielded single scores, although in its later versions the AGCT had separate scores for reading and vocabulary, arithmetic computations, arithmetic reasoning, and spatial relations. Some of the armed services developed and validated tests to measure special aptitudes. Especially in the Army Air Force, under the direction of Flanagan (1948), Thurstone's ideas about primary mental abilities, and his methods of factor analysis, were utilized to develop batteries of special tests for the selection of pilots, navigators, bombardiers, and other classes of personnel. Many of these tests were "printed classification tests" (Guilford, 1947); others were tests involving various kinds of apparatus, motion picture films, and the like to measure perceptual and psychomotor skills. Also, in this and other testing programs much was learned about improved and more efficient techniques of constructing tests, analyzing results, and using those results in personnel decisions (Thorndike, 1949).

Thorndike and Hagen (1959) followed up the subsequent careers of over 10,000 men who had taken Air Force tests during the war, finding that although a factor of general ability differentiated the average scores of people going into different occupations, the predictive value of test profiles for success in those occupations was not impressive.

After the war, J. P. Guilford, who had been in charge of much of the factor-analytic investigation done in the Army Air Forces, built upon this work to conduct, with the support of the Office of Naval Research, a 20-year "high level aptitudes" project to study the full range of mental abilities that might be measured (Guilford & Hoepfner, 1971). In the course of this work he conceived a new model of the structure of mental abilities, which he called "the Structure of Intellect" (SI) (Guilford, 1956, 1967). He and one of his students also published a multifactor battery called the *Guilford-Zimmerman Aptitude Survey* (Guilford & Zimmerman, 1947-1950).

Another important testing and research program that stemmed from work done during the war was Project Talent (Flanagan, Dailey, Shaycoft, Gorham, Orr, & Goldberg, 1962). Inspired by the promising results of factor-analytic investigations he had supervised in his wartime work, Flanagan (1951-1960) had constructed a multifactor test battery called the *Flanagan Aptitude Classification Tests* (FACT). This battery was offered for general use through a commercial publisher, but Flanagan also felt that a parallel battery should be constructed and placed in the public domain so that long-range studies of the identification and utilization of human abilities could be conducted. Flanagan initiated the first major steps for the undertaking in what he called Project Talent. Some 23 aptitude and achievement tests, as well as several questionnaires and interest inventories, were administered to about 140,000 high-school students around the country in March 1960. There have already been several long-term follow-ups of the students in this sample, disclosing important new information about the growth of mental abilities and the way in which they interact with students' educational and occupational choices (Flanagan & Cooley, 1966; Flanagan & Jung, 1971; Jencks et al., 1972).

Multiple-factor analysis theory was also brought to bear in the construction by the United States Employment Service (USES) of an extensive series of tests to measure a number of special mental and other abilities thought to be important in different occupations (Dvorak, 1947). The USES published reports about the validity of these tests for predicting success in numerous occupations.

Commercial publishers were alert to the possibilities of developing multifactor batteries for wide use in secondary schools. One of the most highly regarded of such batteries is the so-called DAT, *Differential Aptitude Tests*, published by the Psychological Corporation (Bennett, Seashore, & Wesman, 1947-1969). This battery offers separate scores on Verbal

Reasoning, Numerical Ability, Abstract Reasoning, Space Relations, Mechanical Reasoning, Clerical Speed and Accuracy, and Language Usage (with separate scores on spelling and "correct" English usage). Little attempt was made to make the subtests "pure" measures of primary mental abilities; the goal was to develop tests that would have the greatest practical use in educational and vocational counseling in high schools. Some of the tests emphasize generalized mental abilities; others stress mastery of skills and knowledges taught in schools. Much labor was expended in item analysis, standardization, and validation of the tests; voluminous data are presented in accompanying manuals concerning correlations of test scores with success in various subjects, with college grades, and with educational and vocational placement after graduation from high school. The "differential" aspect of this battery implies that differences among scores on the several subtests are intended to be useful in counseling; the publishers have provided a casebook that illustrates how counselors can employ profiles of test scores. (Below, this test is examined in more detail.)

Many other multifactor batteries were published during the period, and continuing interest in and demand for them is suggested by the fact that several entirely new batteries of this type have recently come to my attention, for example, the *International Primary Factors Battery* (W. O. Horn, 1973) and the *Comprehensive Abilities Battery* (Hakstian & Cattell, 1976), based on the continuing research (reviewed by J. L. Horn, 1976) that has been done in the identification of factors of mental ability.

In the meantime, there was movement toward the much greater use of mental ability tests in the college admissions process, and there was more extensive research into the refinement of such tests. We have already given brief mention, in the overview of the early period of the test movement, to the work of Carl C. Brigham in the experimental development of a so-called scholastic aptitude examination. It is appropriate at this point to describe this work in more detail, because its effects were felt mainly in the modern period of mental testing. Brigham's early experiences, as described in an interesting autobiographical memoir by Downey (1961), included writing a doctoral dissertation on the Binet intelligence tests and service under Yerkes in the development of the Army psychological examinations in World War I. At Princeton in the 1920s, he developed a test that was used as early as 1925 in connection with admission to the university. Shortly thereafter, the College Entrance Examination Board turned to him to construct a similar examination for experimental use in a small number of "Ivy League" colleges. One of Brigham's innovations at that time was the use of "experimental sections" in the tests of successive years to explore the usefulness of new items and item types and to serve as a basis for making successive forms relatively equivalent in their derived scales. The intensive study that Brigham devoted to each

new item of his tests involved tabulations of answers to each alternative of every multiple-choice item and the use of a mechanical apparatus for timing students' responses to items. Brigham was evidently the first to develop systematic item-analysis procedures using large samples. He was also the first to emphasize the importance of "test security" (accounting for the return of all examination materials) in the administration of admissions tests.

Brigham carefully avoided calling his test an "intelligence test." Although he had earlier believed that tests like the Army Alpha measured intelligence and that the low average scores of certain immigrant populations indicated poor prospects for their success in America, he retracted these views in a famous article published in 1930 (Brigham, 1930). Downey (1931) gives this quotation from one of Brigham's private memoranda:

The testing movement came to this country some twenty-five years ago accompanied by one of the most serious fallacies in the history of science, namely, that the tests measured *native intelligence* purely and simply without regard to training or schooling. I hope nobody believes that now. The test scores very definitely are a composite including schooling, family background, familiarity with English, and everything else, relevant and irrelevant. The "*native intelligence*" hypothesis is dead. (p. 27)

Brigham also recognized the important difference between verbal and mathematical aptitudes (as revealed in the relatively low intercorrelations between tests of these aptitudes). The Scholastic Aptitude Examinations (SAT), first adopted by the CEEB as a regular part of its admissions examination program in 1937, have to this day included separate verbal and mathematical sections, each yielding a score on a scale that Brigham himself introduced—a scale with a mean of 500 and a standard deviation of 100 and with a range of 200 to 800, in a defined population of college admissions candidates.

Under Brigham's direction, the CEEB built a staff in Princeton, near the university, to construct the successive editions of the SAT (and later other objective-type CEEB examinations). This staff carried on after Brigham's death in 1943, continuing to meet Brigham's high standards of test construction. For various reasons that need not be gone into here, this staff was merged in 1947 into a new organization, the Educational Testing Service, that had been proposed by James B. Conant, then president of Harvard University. This organization undertook the further development and administration of the CEEB examinations, including the SAT, as well as the construction and administration of many other testing programs, many of which have involved tests of general intellectual development (e.g., the aptitude tests of the Graduate Record Examination and the National Teacher Examination). Educational Testing Service also became a major center of research on the theory of mental tests, with the

addition of persons such as Harold Gulliksen, Ledyard Tucker, Frederic Lord, and William Angoff to its staff. Gulliksen and Tucker, it may be noted, had been close associates of Thurstone at Chicago. Through the research of these specialists in mental test theory, the measurement characteristics of tests produced by ETS have been of particular excellence, although it cannot be said that *all* their research has yet been reflected in the actual development of tests.

Throughout the years of the modern period, the SAT gained wide acceptance not only among the Ivy League colleges that constituted the original membership of the CEEB, but also among numerous state universities, engineering schools, and public and private colleges in different parts of the country. Whereas, for purposes of studying candidates' qualifications, the early constituents of the CEEB were more interested in the results of essay achievement examinations in various subjects, these and other CEEB constituents increasingly found more usefulness in the SAT scores, as opposed to scores on achievement examinations. (The essay examinations of the earlier days were for the most part abandoned after 1942.) SAT scores, many admissions officers concluded, were better indicators of the true promise of many capable students from unknown high schools who may not have been as well prepared (or coached) as their preparatory-school competitors. In this way, use of SAT scores in the college admissions process tended to even out the disparities in educational advantages between candidates from the wealthier, upper-class families and those from more plebeian backgrounds.

The SAT did not continue to be alone in its field. In 1959, E. F. Lindquist of the University of Iowa, having resigned from his associations with the CEEB (he had been a member of the Board's Committee of Aptitude Examiners that monitored the construction and administration of the SAT), established the American College Testing Program, which quickly grew to serve large numbers of colleges and universities, particularly in the Midwest and West. In contrast to the SAT, the ACT battery of examinations is much more heavily loaded with materials measuring generalized educational achievements in English usage, mathematics, social studies, and natural science. Above all, it seems to stress *reading* comprehension ability. According to a reviewer, "ACT suffers by comparison with the SAT in psychometric care and sophistication, is about equal in validity for predicting collegiate success, and excels somewhat in the variety and extent of ancillary services offered" (W. L. Wallace, in *Buros*, 1972, p. 615).

2.2 *Psychometrics and Theories of Intellectual Abilities*

To provide a perspective from which we can view modern developments in psychometrics, it seems desirable at this point to examine the logical foundations of a viable theory of intellectual abilities. To do so, it is

necessary, unfortunately, to couch the discussion in rather general and abstract terms. Also, brevity demands that many essential details be omitted and many terms be left inadequately defined.

Mental tests typically consist of collections of tasks that examinees can be asked to perform in some given (relatively short) period of time. These tasks, or at least some components of them, may be considered to be a sample of the almost infinite domain of tasks that an individual might be required to perform in some conceivable context in the course of real life, but the samples of tasks appearing in mental tests are constrained in at least two important ways.

First, from a practical point of view, only certain kinds of tasks can conveniently be included in a psychological test. They must be tasks that are small enough to be completed successfully by at least some examinees in a relatively short period of time, and normally they must not involve any complicated apparatus for stimulus presentation or for the examinee's performance of his response. This constraint may appear to make the tasks somewhat artificial; for example, it is not normal in real life for an individual to be confronted with the task (such as often appears in mental tests) of identifying which of a group of words does not fall into the same semantic classification as the others of a set. Nevertheless, it might be argued that such "artificial" tasks represent or contain elements of larger tasks that might be encountered in real life.

Second, the tasks in a mental test are selected in such a way that it is reasonable to suppose that (except for "lucky guesses") successful performance of a task requires some one or more kinds of intellectual competence or potential and that any other kinds of competence that might be required are not critical, in the sense that all examinees may be assumed to possess these noncritical competences in a sufficient degree. (For example, it is normally assumed that all examinees taking a "paper-and-pencil" test can in fact hold a pencil and make certain required kinds of marks with it.) Initially, the selection of tasks for a mental test is a matter that is up to the judgment of the test constructor, but there are, as we shall see, empirical operations whereby the test constructor's judgments become less critical.

Given the characteristics of mental tests as we have described them, we now attempt to state the requirements of an adequate theory of such tests, that is, a theory of what such tests measure and how well they measure whatever that may be. An adequate theory, it would seem, ought to give a satisfactory account of at least the following:

1. *The extent to which the tasks, singly or in groups, can or do yield reliable measures of the characteristic behaviors of individuals performing such tasks.* To the extent that the measures derived from the tasks are in fact reliable (and we will leave this term undefined for the moment), it may be inferred that individuals may be arranged or ordered on a scale or

continuum reflecting different degrees of competence with respect to the class or classes of tasks represented by those for which their performance has been observed.

2. *The extent to which successful performance on each task, or group of tasks, reflects some particular kind of cognitive competence or potential.* The theory must therefore offer a satisfactory classification of the various kinds of cognitive competences that might be observed. To each kind of competence it may be assumed that there corresponds a scale or continuum containing different degrees of competence.

3. *The source or cause of the individual differences observed with respect to any scale of cognitive competence so identified.*

These three requirements of a theory of intellectual abilities constitute problems to be solved empirically. Psychometric theory is concerned primarily with the first two of these problems. What is conventionally termed "test theory" is addressed to the first problem, and what has come to be known as "factor analysis" is devoted to the second problem. The third of these problems can in principle be approached only by appeal to more general methodologies of investigation in the behavioral sciences, for example, by observational and experimental studies, and will not be dealt with here except in passing. Psychometric techniques, nevertheless, play a critical role in this third class of studies because they are essential in establishing reliable and distinguishable variables for use in such studies.

2.3 Test Theory in the Modern Period

As we have seen, the foundations of test theory were laid in the early period of the mental test movement, initially by Spearman, Brown, Thorndike, Kelley, and others. Test theory in the early period, however, was concerned chiefly with the "reliability" of measurements that were *already quantified* (e.g., "number correct" scores on mental tests, and "intelligence quotients"); concern for exactly how this quantification took place was expressed only in the development of various "item analysis" techniques that would ensure that all items or tasks involved in obtaining a given measurement would tend, on the average, to be consistent with each other and thus contribute effectively to the reliability of the measurement. Reliability was defined in terms of the extent to which observed measurements reflected a "true" underlying score as opposed to "error." This approach has come to be known as "classical" test theory (Novick, 1966; Stanley, 1971).

In the modern period, there have been many elaborations of test analysis techniques arising from classical test theory. Gulliksen's text (1950), Lord and Novick's (1968) treatise, and Stanley's (1971) chapter summarize many of these. Kuder and Richardson's (1937) formulation of a class of procedures for estimating reliability from the "internal consis-

tency" of the items or tasks composing a test and Cronbach's (1951) derivation of a so-called coefficient alpha, a generalization of some of the Kuder-Richardson results, led to determinations of reliability that were in some ways more accurate (because they took more data into account) and more convenient (because they could at the same time be based on the administration of a single test). Test theorists began to look more closely at the effects of different distributions of item difficulties on the measurement characteristics of total test scores. Over a considerable period, attention was devoted to the so-called attenuation paradox (Tucker, 1946; Loevinger, 1954; Lord, 1955) whereby it seemed that maximum reliability, but reduced validity in terms of measurement of an underlying ability, was attained when the distribution of item difficulties was "peaked," that is, when all item difficulties were near 50% difficulty. The paradox was resolved when it was shown that it depended on misapplication of correlation coefficient formulas and that in many circumstances the distribution of item difficulties should not be peaked at all. (The issue has, in fact, led to a great many misunderstandings among test analysts, misunderstandings that prevail even now in some quarters.) A related issue that has been studied extensively is that of how item scores can be weighted to produce optimal measurement characteristics of total scores (Stanley & Wang, 1970). In general, special weighting schemes have little advantage over simply counting items with unit weights.

A current concern of test theorists that is essentially an outgrowth of classical test theory is the development of what Lord (1971) has called "tailored testing," that is, a procedure for optimizing the efficiency of a mental test by the sequential administration of items. In early stages of such testing, a brief test composed of items of a wide range of difficulty is administered to make a rough determination of the examinee's level of ability; the subsequent items are selected in such a way as to "zero in" on the level of ability so established. The motivation for developing these procedures seems to have arisen from the possibility of administering tests by an interactive computer terminal; extensive efforts by the College Board to adapt such procedures to group testing resulted in the finding that they resulted in very little significant improvement in the accuracy of measurement and were rather impractical besides. It is ironic that tailored testing is in a sense a modern adaptation of procedures that have long been used in connection with individual tests such as the Stanford-Binet.

Other modern developments from classical test theory have included consideration of the implications of unreliability for the measurement of change or test gains (Cronbach & Furby, 1970), the development of optimal methods for establishing equivalent scales on different tests of the same trait (Angoff, 1971b), and studies of various sources of unreliability such as sampling of items, and the effects of practice, fatigue, and chance guessing.

Contrasting with the "classical" approach to reliability is one that emphasizes problems of statistical estimation and that is derived from the work of R. A. Fisher (1925) in statistical inference. Seminal papers by Baker (1939) and Hoyt (1941) got this "ANOVA" (analysis of variance) approach off the ground; later developments by Stanley (1961) and others have made it possible to determine the relative effects of different sources of variation—underlying ability as opposed to variation due to "test form," task difficulty, etc., and their interactions. Stanley (1971) believes that this ANOVA or "components of variance" approach makes for much greater precision and flexibility in the estimation of reliability. A still further development of this approach, called "generalizability theory," is due mainly to Cronbach and his associates (Cronbach, Rajaratnam, & Gleser, 1963; Cronbach, Gleser, Nanda, & Rajaratnam, 1972), providing procedures for estimating the reliability of a test in terms of the degree to which the scores represent those that would be derived from a "universe" of measurements (items, raters, or the like).

A persistent theme in the development of mental tests is a concern with the way in which successful versus unsuccessful performance on a given item or task relates to an underlying "latent trait" continuum of ability. Binet (Binet & Simon, 1905) assumed that a one-to-one correspondence could be established between graded tasks and mental ages, but experience with Binet-type scales revealed that the correspondence was far from exact. What is needed, either in connection with Binet-type scales or in connection with the "point-scales" associated with group mental ability tests that are scored in terms of number correct, is a model for the probability of an examinee's successfully passing an item or task as a function of (among other things) the characteristics of the task (principally, its "difficulty" and "discriminating power") and the examinee's position on an underlying continuum of ability. Such a function has been called the item-characteristic function (Lord & Novick, 1968). A satisfactory model of the item characteristic function should enable one to establish units of measurement for the ability scale, demonstrate the extent to which two or more items measure this continuum to known degrees, and derive information on the examinee's placement on the ability scale from his performances on the set of items thus selected.

Progress in developing satisfactory models of this type has been relatively slow; for example, Ferguson's (1942) and Finney's (1944) contributions, in which they assumed that item success is related to the ability scale by a cumulative normal distribution function, represented only a slight advance over that of Thurstone (1925b) 17 years earlier. Modern developments are due primarily to Lord (1952; cf. Lord & Novick, 1968), Birnbaum (1968), and Rasch (1960), but these developments are really only elaborations of the earlier models, taking more parameters into account or resorting to the use of somewhat more tractable mathematical

expressions. Their models do, however, provide reasonably satisfactory solutions to some of the problems of test construction, except that they are limited to circumstances in which one can assume that every item measures the same ability (or composite of abilities).

Satisfactory models for determining whether two or more items measure the same ability (or composite of abilities) and for relating performances on such items to underlying scales of ability have been elusive and controversial. Loevinger (1947) took the tack of contrasting a model of a "perfectly homogeneous test" (in which all items measure an underlying scale perfectly) with that of a completely "heterogeneous" test in which all items measure different abilities (each perfectly, however). Her "perfectly homogeneous" test actually followed the model described earlier by Walker (1931), Guttman (1941), and Carroll (1945) in which if an examinee passes an item of a given level of difficulty, he passes all items of lower levels of difficulty, and, conversely, if an examinee fails an item of a given level of difficulty—he also fails all items of greater difficulty, difficulty being defined in terms of the proportion of examinees failing an item. Loevinger's contribution was to suggest a measure of the extent to which a test approximates perfect homogeneity, in her sense of the term. There are, however, a number of still unresolved difficulties in Loevinger's approach, since it does not take systematic account of sources of unreliability; consequently, most tests show relatively low degrees of homogeneity—probably spuriously low.

It would seem that the question of whether sets of items can be regarded as measuring the same abilities could be answered by techniques of correlational or factor analysis, but these techniques present difficulties stemming from the inappropriateness of the usual methods of computing correlations (Carroll, 1961). Some of these difficulties are overcome by methods proposed by Wherry and Winer (1953) and Loevinger, Gleser, and DuBois (1953), among others. Few improvements have been made in these methods, however, and they have not been widely applied in test construction. In this connection, it should be noted that Wherry and Gaylord (1943) were the first to point out that the interpretation of reliability coefficients determined by Kuder-Richardson, or similar methods, required the (often false) assumption that all the items measure a single ability (or composite of abilities). This fact has often been ignored; Lord and Novick (1968, p. 95) dodge the issue by observing that there is no general agreement as to the meaning of homogeneity. They nevertheless offer reasons for concluding that a perfectly homogeneous test is one in which Cronbach's coefficient alpha is equal to the theoretical reliability of the test; the problem here is that there is no way of confidently determining the theoretical reliability of the test.

In the state of current knowledge, there is no satisfactory way of guaranteeing the construction of tests that are completely "homogeneous"

in the sense that all items measure the same ability despite the presence of unreliability of measurement.

Several recent discussions have expressed considerable dissatisfaction with the present status of mental test theory. Lumsden (1976), for example, has this to say:

Test theory has had few major ideas. I can list only five.

1. the decomposition of obtained scores into true and error components;
2. the duality of psychophysics and psychometrics;
3. the notion of unidimensionality;
4. the conception of test validity as theoretical equivalence, usually called construct validity;
5. the scaling ideas derived from various item characteristic curve models.

Apart from the first, these ideas have been hesitatingly and unconfidently applied. (p. 251)

Lumsden gives less credence to Bock and Wood's (1971) suggestion that test theorists are discouraged by the uncritical use of tests than to his own belief that "the shreds of theory that have been developed and the timeworn true score models are not rich sources of ideas about testing, and the ideas that are generated seem to relate mainly to the internal workings of the models." Likewise, Levy (1973) comments that "the holy cows of item analysis, standardization and indices of reliability appear to receive more attention from test constructors than elucidation of meaning" and urges that we need "tests constructed to test hypotheses, and fewer hypotheses about tests" (p. 36). Despite their acid criticisms, both of these writers strike a somewhat optimistic note by giving examples of promising test development activities, or prescriptions for them, like Lumsden's (1976):

The new test theorists . . . will be primarily test constructors and validity people who attempt to realize their dreams. . . . They will not test a new model with a few items from the SAT files (or from a computer), find a mediocre fit to some dubiously relevant criterion, and then go on to the next. Rather they will set out the requirements for the application of the model against a user-for-blood standard of efficiency. (p. 277)

While mental test theory has made considerable progress toward establishing a satisfactory theory of mental test scores, it appears that it has made little progress towards a theory of the abilities that those scores presumably measure. We turn now to consider whether any progress toward such a theory has been made by a different and by now quite separate branch of psychometrics, factor analysis.

2.4 Factor Analysis in the Modern Period

Introducing his review of "factor analysis to 1940," Dael Wolfle (1940) remarked:

A single correlation coefficient is relatively easy to interpret. If the correlation is high, it may be assumed that there is considerable overlap between the determiners of the goodness of performance; if the correlation is low, it may be assumed that there is little overlap between the two sets of determiners. But if one is faced with a large table of correlations, the task of explaining all the complex interrelations is almost hopeless. Factor analysis provides one method of summarizing these relations, thereby making it easier for the psychologist to interpret and explain them. (p. 1)

Wolfe went on to state two important goals of all factor analysis investigations: simplification—explaining the correlations by means of a small set of values that can be regarded as having generated the correlations; and the association of these values with a small set of “factors” that will be “more fundamental in nature than the original tests.”

Considerable progress in factor analysis had been made in the early period; the main achievement had been Spearman's elaboration of a theory that claimed that correlations among mental tests could be explained chiefly by one common or “general” factor and that any residuals were to be attributed to the influence of “specific” factors. It proved impossible to support this claim in all cases, however; various mathematical techniques were advanced, chiefly in Great Britain, to show that correlation matrices often contained evidence of the influence of “group” factors in certain sets of tests. The methods developed to analyze correlation matrices for group factors were somewhat subjective and cumbersome, however.

During the 1930s, L. L. Thurstone, at the University of Chicago, developed methods of factor analysis that made no initial assumptions concerning a “general” factor or “group factors”; they were intended to arrive at the most parsimonious and psychologically meaningful description of the data to which they were applied, whether the data contained one, two, or any number of factors. Some of the important ideas advanced by Thurstone in an extensive series of publications (Thurstone, 1931b, 1935, 1940, 1947) were the following: the notion of *common factor variance*, and the associated *communalities* of particular tests, representing the variance that exists in common among the tests in a particular set and thus the variance that can be analyzed into common factors; the notion of *rotation to simple structure*, whereby a purely mathematical description of a set of data was transformed to a psychologically meaningful and (in some ways) more parsimonious description; and the notion of *correlated factors* and *oblique factor structure*, whereby underlying entities or “factors” could themselves be correlated and explained by a further factoring at a higher order of analysis. Thurstone also introduced several mathematical techniques of factoring that were computationally simpler, though less elegant, than some of those advanced by Hotelling (1933), Burt (1938); and others; chief of these was the so-called centroid method. (Nowadays

the more elegant methods are usually used, because the high-speed electronic computer has made them practicable and convenient.)

The fundamental model underlying any method of factor analysis, as Thurstone saw it, is one whereby any observed test score of any examinee is to be explained or interpreted as a weighted sum of the examinee's scores on a set of underlying hypothetical entities called factors or, sometimes, latent traits. Generally there is interest only in so-called common factors, that is, factors that can be shown to manifest themselves significantly in performance on more than one test variable. (Any factors that remain specific to given test variables are attributed to error or to the idiosyncrasies of the particular test variables.) The model whereby a test score is a weighted sum of scores on underlying factors is, in statistical terms, a "linear" model similar to the kind of model used in many other branches of statistics. Such a model implies, incidentally, that high values on one underlying factor can compensate for low values on another underlying factor. Two individuals having the same observed score on a given test variable could therefore be assumed to arrive at that score in quite different ways, one by being very good in one ability and poor in another ability, the other by having the converse pattern of abilities. Although nonlinear models of factor analysis are now available (McDonald, 1962), there is as yet no evidence of serious difficulties in accepting the implications of the purely linear model for the usual types of ability tests.

Since one has no way of knowing in advance what the underlying entities or factors are, or how they operate to generate observed data, one must try to solve for the parameters of a factor analysis model by applying one of various approaches to the analysis of empirical data from a set of variables. Usually the data analyzed are in the form of a matrix of correlations among the variables. The problem of factor analysis, therefore, may be thought of as one of determining how many common factors are needed to explain the correlations among a set of variables and then determining an appropriate and psychologically meaningful set of weights for these variables. Thurstone's methods provided, by the mid-1930s, what was widely regarded as a more satisfactory and reasonable group of procedures for solving this problem than had been previously available, and these methods were adopted in most of the factor-analytic research in America.

From today's perspective, Thurstone's methods yield results that are generally quite similar to those yielded by procedures developed elsewhere, but perhaps a principal reason for the acceptance of Thurstone's methods was the fact that the first major application of his procedures produced strikingly interesting results. Thurstone's skill in test construction also contributed to his success. His monograph *Primary Mental Abilities* (1938a) was a factor analysis of 57 assorted mental test

variables, yielding 13 centroid factors that were rotated to a "simple structure." Nine of the factors were interpreted psychologically (some quite definitively, others somewhat tentatively) as representing distinct types of mental abilities:

- S - Space
- V - Verbal Comprehension
- W - Word Fluency
- N - Number Facility
- I - Induction
- P - Perceptual Speed
- D - Deduction
- M - Rote Memory
- R - Reasoning

Subsequent analyses (Kaiser, 1960) of these same data, using more advanced methodologies, have shown that most of the original factors stand, although there have been slight reinterpretations of the meanings of the factors. In the framework of the factor model, each of these factors was thought of as a distinct underlying ability that would manifest itself in any test that called for the use of that ability. Individuals who had a low level of a particular ability would be handicapped in their performance on any test that involved that ability, particularly if the test were a "pure" test of that ability in the sense that it called on no other abilities.

Thurstone's results lent support to a rather different concept of intelligence from the one that had dominated the earlier period of the test movement; that is, they supported the concept that there is not *one* kind of intelligence, but *many*. Most of the tests that Thurstone used in his PMA study were expanded forms of tasks or items that had been used in the conventional group intelligence tests of the period. From the factor-analytic viewpoint, group intelligence tests were "hodgepodes" of mental tasks; the single scores derived from such tests were seen as masking the true pattern of abilities that an individual might have. If an individual received a high score on one of these "omnibus" intelligence tests, it could mean simply that he happened to have a high standing on all or nearly all of the abilities tapped by such a test. An average score could reflect any one of many different patterns of abilities—either average standings on a number of abilities or some combination of high and low standings on different abilities.

Further, the new results opened the way to the idea that in the realm of the intellect, people could have a wide variety of strengths as well as weaknesses. This fitted in with the notion that people might differ widely in their aptitudes for different educational programs, occupations, and avocations. A child with a relatively low standing on general intelligence

was not thereby doomed to failure in everything of an intellectual nature that he might undertake; rather, it was thought, his low score on the composite test might conceal some particular kind of talent.

In the early years of factor analysis investigations, there had not been time to assemble evidence as to whether the underlying factors were of genetic or of environmental origin, or from some combination of these influences. The common opinion, however (possibly influenced by the intellectual climate of the time), seems to have been that the "factors" identified in factor-analytic investigations were largely of genetic origin.

From the standpoint of the requirements for an adequate theory of intelligence tests, the significance of multiple-factor analysis methodology was that it seemed to offer an objective way of identifying and classifying the different kinds of competence that might be found in test performances.

In the modern period up to the present, factor-analytic research has proceeded along two relatively independent lines. On the one hand, efforts have been directed toward the improvement and refinement of the methodology itself; on the other hand, the available methodology has been widely applied in the search for new factors of ability and more satisfactory psychological interpretations of the factors identified.

Progress in methodology. Despite his extensive contributions to factor-analytic methodology, Thurstone was able to offer only partial or approximate solutions for several difficult mathematical and theoretical problems. Mainly, these are problems of indeterminacy. One of these concerns how many factors one should accept as being significantly in the "common-factor space." A related problem is how best to estimate the "communalities" of the variables, that is, what proportion of variance in each variable is to be identified as being in the common-factor space. These problems arise principally because empirical data always contain at least some error variance, and thus it is difficult to decide where to draw the line between true common-factor variance and error variance.

A second major problem has to do with the notion of "simple structure" and the rotation of the coordinate framework to yield a structural description of the data that is regarded as psychologically meaningful and interpretable. This problem arises from the mathematical fact that there is an infinity of factor matrices that can generate a given correlation matrix, analogous to the familiar fact that there is an infinity of pairs of numbers whose product is equal to some given number. Thus there is a problem of selecting the matrix that is most readily interpretable.

A third major problem has to do with the kinds of models that one tries to fit to factor-analytic data—either models that are maximally parsimonious in terms of the number of underlying factors hypothesized or models that although they are less parsimonious may provide a more

reasonable description of the data from a psychological point of view.

Considerable progress has been made toward solving all these problems, which are interrelated in complex ways.

Among major contributions to the first problem (estimating the number of significant common factors and thereby estimating the communalities of test variables) have been Guttman's (1940) proof of a lower bound to the communalities and his development of image factor analysis (1953); Kaiser and Caffrey's (1965) development of "alpha factor analysis," relating factor analysis to the reliability and generalizability concepts of test theory; and Jöreskog and Lawley's (1968) further development of methodologies that permit significance tests. These last authors introduced a distinction between "exploratory" factor analysis, in which one simply seeks statistically sound factorial descriptions of data, and "confirmatory" factor analysis, in which one tests statistical hypotheses about the structure of the data.

The second major problem—that of rotation to simple structure—has been approached by so-called analytical methods that rely on some type of mathematical criterion of good fit, replacing the subjectively determined graphical rotation methods employed by Thurstone and others. Analytical methods of rotation were first introduced in 1953 (Carroll, 1953; Saunders, 1953), but the currently most widely used and generally satisfactory procedure is that due to Kaiser (1958), the so-called varimax method for orthogonal factors. There now exist at least a half-dozen analytic methods from which a researcher can select according to his taste; in general, these methods yield highly similar results for a well-designed study. The problem of indeterminacy in rotation of factor axes is seen to lie essentially in the adequacy or inadequacy of the data subjected to factor analysis.

There has been less interest in the investigation of different models of factor analysis. One of the major contributions falling in this category, however, is Schmid and Leiman's (1957) development of an objective method of determining, from Thurstone-type results, a hierarchical factor structure that is not as parsimonious as Thurstone's but that conforms more closely to the factor models preferred by the British school of factor analysis (with general, group, and narrow factors). This method, which has much to recommend it, has not been applied as widely as one might desire. Other developments in this category include McDonald's (1962) nonlinear factor analysis model and Tucker's (1963) three-mode factor model through which one can study factor-analytic data with an added data dimension such as different testing occasions, different testing methods, or different observers. The utility of these methods has not as yet been sufficiently explored.

A general comment on the present state of the art of factor analysis is that even if there is still some room for further refinement, methodologies

are now available to do about anything one might like to do with factor analysis. The advent of high-speed computers, beginning around 1954 (as far as psychometricians were concerned), was particularly critical for the development of factor analysis because of the large and complex computations required. Computations that would have taken months in 1940, if they were thought practicable at all, can now be performed in a matter of seconds. Perhaps the only aspect of factor analysis that is in great need of further development is the factor analysis of large numbers of binary test items according to appropriate scaling models.

2.5 *Factor-analytic Studies in Cognitive Abilities*

Thurstone himself, together with his colleagues and students, followed up the findings of his Primary Mental Abilities study (1938a) with an extensive series of investigations of various factor "domains," such as those of perception (Thurstone, 1944), language (Carroll, 1941; Taylor, 1947), and number (Coombs, 1941). The object was to confirm or to refine the interpretations of the factors discovered in the *PMA* study and to see whether further factors could be disclosed. Systematic variations were introduced into the variables in order to test hypotheses suggested by ambiguities in the interpretation of previous results. For example, it was found that facility in numerical computations was best tested when the computations were extremely simple and speeded, as in the addition or multiplication of pairs of single-digit numbers; Coombs also obtained results that suggested that the "Number" factor was not limited to numerical material but could also operate in the manipulation of other well-learned symbolic rule systems.

The difficult and ambitious enterprise of studying cognitive abilities by factor-analytic methods has been carried on since Thurstone's day by numerous investigators, not only throughout the United States, but also in such countries as Sweden (Werdelin, 1961), Switzerland (Meili, 1964), and Germany (Jäger, 1967). By 1951, French (1951) compiled a summary of factorial investigations in which he listed 59 different factors of ability and scholastic achievement that had been "identified with sufficient certainty to receive a name." French and his colleagues at Educational Testing Service assembled two successive "kits" of tests of confirmed factors of mental ability (French, 1954; French, Ekstrom, & Price, 1963). In preparation for making a third such kit, Ekstrom (1973) published a review of the existing literature on cognitive factors, concluding that there was sound evidence for at least 20 different factors of mental ability; the resulting kit (Harman, Ekstrom, & French, 1976) has tests of 23 factors.

One of the most active investigators has been Guilford, who directed a 20-year "high-level aptitudes" project and who by 1971 (Guilford & Hoepfner, 1971) claimed the confirmation of some 98 factors of cognitive

ability. Other active investigators include Cattell (1971) and Horn (1968), whose concern with the relations of cognitive ability factors to personality traits sets them somewhat apart from other groups of investigators.

The proliferation of mental factors beyond the small number—about seven—identified with reasonable confidence by Thurstone up to the large numbers reported by Guilford and his colleagues has been puzzling to many psychologists. Does this proliferation result somehow from artifacts of the methodology, or does it correspond to an actual complexity of "mental life"? Questions have also been raised about how many factors are of substantial importance in the conduct of everyday life, education, and other activities; it had been observed (Ellison & Edgerton, 1941) that relatively few of Thurstone's primary mental abilities had significant correlations with any kind of scholastic success. (Chiefly, only verbal and reasoning abilities were found to have more or less reliable correlations with school grades.)

Responses to the "proliferation" problem took several different directions. Methodologists, of course, retreated to their studies and computer terminals to reexamine the statistical bases for establishing factors (e.g., Linn, 1968). Guilford, however, saw the proliferation of factors as an indication that mental life is structured in such a way that it is entirely reasonable to expect the existence of a large number of factors. During the course of his 20-year aptitude project, he developed a three-dimensional taxonomic model for classifying the large number of factors he claimed to have discovered. This model, which he called the Structure of Intellect (SI) model, postulated that a given factor could be classified in terms of the Contents that were involved, the mental Operations that were required in tests of the factor, and the types of Products that resulted from these operations. The model set forth four types of Contents, five types of Operations, and six types of Products. On the assumption that a factor could be defined in the terms of exactly one type of Content, one type of Operation, and one type of Product, and that any combination of Content, Operation, and Product could generate a factor, Guilford predicted that continued research would eventually be able to identify 120 different factors. Having found factors representing a great many of these combinations, he directed his efforts toward finding factors that represented all of the remaining combinations in his taxonomic model. Guilford's SI model has been criticized on various grounds, partly methodological (Horn, 1967) and partly logical (Carroll, 1972). The methodological criticism was that Guilford was too ready to accept factors on the basis of subjective judgments and fits to his data that might actually be due to the operation of chance. The criticism from the standpoint of logic was that there was nothing in Guilford's data that *required* or even positively indicated the classificatory constructs that he postulated, since

all the factors are, in any event, claimed to be orthogonal and independent.

A different approach to the problem was implicit in the work of Cattell (1971). Some years ago, Cattell (1943) had called attention to a possible distinction between what he called "fluid" intelligence and "crystallized" intelligence, the former representing basic capacity and the latter representing abilities acquired through learning, practice, and exposure to education. These were viewed as group factors standing beside a *g* or general factor. Both fluid and crystallized intelligence could be regarded as having associated with them a larger number of factors representing rather narrow abilities at the level of Thurstone's "primary mental abilities." In effect, Cattell espouses a hierarchical model of intellectual abilities; in a hierarchy, it is more or less inevitable that the terminal branches of the tree will proliferate.

Most researchers in factor analysis, it would appear, assume that factors correspond to rather permanent constitutional traits or attributes of individuals—possibly arising in part from genetic sources. The tests measuring such factors are assumed to elicit the operations of these traits—somewhat in the way that a weight-lifting task calls on the individual's muscular powers. But factors can also be looked upon as manifestations of learned competences. It can be easily demonstrated that if two types of content tend to be learned or practiced together (as opposed to both *not* being learned together), tests of those contents will tend to have a positive correlation, giving rise to an underlying factor. The Canadian psychologist George Ferguson (1956) has put forward an interpretation of factor-analytic results that relies on the notion that learning leads to the acquisition of abilities that can generalize and transfer to various performances. A factor, therefore, reflects the set of performances to which an ability can generalize and transfer. The full consequences of Ferguson's position have never been worked out in application to the variety of factor-analytic results that are now available, although a sketch of such an application to results in the verbal ability domain has been offered by Carroll (1962). Ferguson's theory does not exclude the possibility that individual differences in genetic and constitutional traits set differential limits on individuals' development of their abilities, but it draws attention to the way in which learning processes can be used to account for factor-analytic findings. Further discussions of a "learning-theoretic" interpretation of factors may be found in *Intelligence and Experience* by J. McV. Hunt (1961, 297ff.).

The latest developments in the theory of the abilities measured by mental tests have arisen from the "information processing" approach that has been widely adopted in contemporary cognitive psychology. The performance of any mental task is conceived as involving the simulta-

neous (parallel) or sequential operation of mental processes upon inputs from external stimuli or from internal "memory stores," governed by learned "executive" routines, and resulting in further operations or responses involving either peripheral motor systems or internal memory stores. For example, the response to a multiple-choice vocabulary item appearing on a mental test would involve (among other things) reading the words shown in the item, recognizing templates for these words in an internal memory store, retrieving semantic information for these words, comparing different pieces of semantic information, and selecting a response on the basis of these meaning comparisons. Hunt, Frost, and Lunneborg (1973) assume that such mental processes have individual difference parameters like speed, probability of transfer between different memory stores, capacity of memory stores, and the like, and have attempted to correlate subjects' performances on mental tests with their performances on experimental tasks that are specially designed to permit the derivation of these parameters. They have turned up some interesting and suggestive correlations, but their results have not yet been confirmed by other investigators. In a frankly speculative article, Carroll (1976) has tried to interpret the 24 factors measured by the ETS Kit of Factor Reference Tests (French, Ekstrom, & Price, 1963) in terms of Hunt's information processing model, drawing attention to the kinds of mental operations, strategies, and memory stores that may be involved and to the way in which individual differences in underlying parameters may be revealed in factor test scores. A related development is Sternberg's (1975) procedure for the "componential analysis" of mental tasks like the solution of syllogisms.

This line of work is too recent to be confidently evaluated, but it promises to permit much more rigorous statements about the nature of the cognitive abilities measured by standardized mental tasks. Even then, it will be necessary to investigate in detail the influences giving rise to individual differences in the parameters of mental tasks.

2.6 *Theories of Measurement and Scaling*

Serious attention to providing a theoretical basis for psychological measurement came only in the modern period of the testing movement. Foundations had been laid in mathematics (Hölder, 1901) and in the physical sciences (Bridgman, 1927; Campbell, 1928). Bridgman had stressed the definition of physical measurements (of length, mass, etc.) in terms of the operations employed in obtaining them, and Campbell had set out the requirements for what he called fundamental measurement, the assignment of numerals in terms of equality of additive units, as opposed to measurements derived from them. Building on these foundations, B. O. Smith (1938), an educator, was one of the first to examine the

nature of the scales underlying psychological and educational measurements, suggesting that most such measurements are of a derived rather than a fundamental type, in Campbell's sense. (This suggestion was probably misguided, however; psychological measurements are neither fundamental nor derived.)

The system of scale types proposed by Stevens (1946, 1951) has been most influential in current discussions of measurement theory in psychology. Stevens viewed measurement as the assignment of numerals to objects or events according to rules. He recognized four basic types of measurement scales, in increasing order of "strength": nominal, ordinal, interval, and ratio. (Some writers, like Jones, 1971, object to nominal scales as measurements.) Stevens's scale types are differentiated by what properties of the number system are used in defining them; for example, ordinal scales use only the transitive asymmetric relationships of the number system (e.g., 1 is less than 2, 2 is less than 3, thus 1 is less than 3), whereas ratio scales imply equality of units, equality of ratios (thus, e.g., the ratio of 4 to 2 is the same as the ratio of 8 to 4), and an absolute zero. Stevens claimed that psychological measurements are only rarely on ratio or even interval scales; the scales underlying psychological tests of ability, he believed, are no more than ordinal since they provide only a ranking of the individuals measured. If this claim is taken seriously, it has been argued, it implies that many types of statistics that are ordinarily applied to test-score data (e.g., arithmetic averages, standard deviations, and correlation coefficients) are not strictly appropriate; more valid would be "nonparametric" statistics such as medians, percentiles, and rank-order correlation coefficients.

In recent years, work on the theory of psychological measurement has focused on providing a more rigorous axiomatic formulation (Suppes & Zinnes, 1963; Krantz, Luce, Suppes, & Tversky, 1971), but there has as yet been little use of such formulations to consider the conditions under which psychological measurements could have stronger than ordinal properties. Most test theorists have continued to rely on the assumption, widely made by early workers in the field, that abilities tend to be distributed normally, that is, according to the Gaussian normal probability function; they point out that many traits that are measured in fundamental units (e.g., adult height) tend to follow this distribution. Under this assumption it is asserted that if psychological measurements follow approximately normal distributions, or can be transformed to scales that do so, they may be presumed to be measured on scales with at least interval properties. There are difficulties entailed in this assumption, however, stemming from the fact that an approximately normal distribution can be generated simply from the aggregation of a large number of equiprobable events. Observed normal distributions of mental abilities, therefore, may

represent mathematical artifacts, and the inference that their scales have interval properties may be invalid, if only for this reason.

It seems obvious that further work is desirable to provide an adequate theoretical basis for the scales on which mental abilities are measured, or if necessary, to show that it is impossible to provide such a basis.

2.7 Statistical and Computational Technology

While test theorists and factor analysts devoted themselves to the specific statistical problems that arose in the construction of tests and the identification of factors underlying them, statisticians continued to refine procedures of general utility in behavioral science investigations. These procedures were used (and sometimes misused) in studies that were designed to evaluate applications of mental ability tests or to identify the probable causal antecedents of individual differences revealed by such tests.

The modern period saw a great expansion in the technology of the correlational method, used in general to describe relationships among variables. A problem that often arises in connection with mental ability tests, and indeed in many other circumstances, is that of finding the best way of combining a set of independent variables—such as the subscores of a test battery—to predict the values of one or more dependent or criterion variables. The simplest way of combining variables is to add them up, each with some definite weight or coefficient, but the question arises: What are the best weights? A solution to this problem was reached around the turn of the century by Pearson (1896) and Yule (1897), through what is known as the multiple correlation, or multiple regression, technique. An early application of the technique occurred in Yerkes's (1921) report on Army testing in World War I; extensive use of the method was delayed, however, until efficient algorithms for manual computations were adapted by Wherry (1932) from mathematical work due to Doolittle (1878). It was not until the advent of electronic computers in the 1950s that computation of multiple correlations and associated parameters became really practicable for large numbers of variables. An extension of this technique for predicting optimal combinations of criterion variables, known as canonical correlation, had been developed by Hotelling (1935) but did not come into even moderate use until the days of the electronic computer. A further generalization to any number of sets of variables was provided by Horst (1961).

All these methods concern merely the identification and description of relationships among variables, without necessary implication of causality. When measurements of variables are taken at different times, however, some inferences about causal relations may be attempted, as was shown by Wright (1921); this technique of path analysis, so called, has only recently

been applied to any extent in the analysis of psychological data (Werts & Linn, 1970). A closely related technique, the cross-lagged panel correlation, has been applied, for example, in a study of whether intelligence "causes" achievement (Crano, Kenny, & Campbell, 1972). These and other techniques have greatly increased behavioral scientists' ability to identify and interpret relationships between mental ability tests and other variables (Tatsuoka, 1973).

Much progress has been made in refining other aspects of statistical technology, such as sampling theory and experimental design. Sampling theory can be important in the design of large-scale surveys in which mental test data are involved. The theory of experimental design could conceivably be called into play in the design and analysis of mental testing procedures conducted under very carefully controlled conditions, as in an experimental laboratory.

The great advances in computer technology that began to be made in the late 1940s not only facilitated the development of new procedures in the statistical analysis of psychometric data (Green, 1966; Cooley & Lohnes, 1962) but also were accompanied by corresponding advances in the technology of scoring tests and manipulating test-score data. It became possible to score test answer sheets very rapidly and accurately by optical scanning devices, and to produce elaborate test-score reports by computer. In a number of schools and universities, facilities became available whereby students could take tests at a computer terminal. The possibility of having such tests "tailored" for the particular ability level of the examinees has been under investigation (Lord, 1971), but as yet the practice of giving such tailored tests has not become widespread. Most group tests of intelligence and scholastic aptitude are still administered in traditional ways; that is, examinees, seated at desks, make their answers by marking in test booklets or on answer sheets, and (typically) are given a uniform time limit to complete their work. In this respect, the process of administering group tests has changed very little in the last fifty years. Nevertheless, the kind of research being explored by Hunt, Frost, and Lunneborg (1973; Hunt, 1976) and Sternberg (1975) may suggest new procedures in test administration, involving the extensive use of computer terminals.

2.8 Practices in Test Construction, Standardization, and Validation

As was done in the description of the early developmental period of the testing movement, practices in test construction, standardization, and validation in the modern period can be illustrated through an examination of several widely used group intelligence and scholastic aptitude tests.

The 1973 Henmon-Nelson Tests of Mental Ability. Earlier we examined the 1932-1935 edition of these tests. Except for a few refine-

ments in format, the 1973 edition (French, 1973) is highly similar to earlier versions. The number of items, the arrangement of the items in "cycle-omnibus" form, and the time limits are the same as they were forty years before. The test format and printing are somewhat modernized, and the examinee can make his responses either with the Clapp-Young Self-Marking Device used earlier or on an "MRC" card (the same size as a standard Hollerith/IBM card, but printed with spaces for answers) so that the test answers can be scored at a central point (i.e., the Measurement Research Center at the University of Iowa). The technology of test construction has been brought to bear on the test only with respect to the selection of items and the procedures of norming and validation. Actually, the *types* of items are very much the same as in earlier editions. The test manual (French, 1973) does not attempt to describe the processes by which new items were generated, but it states:

In selecting items for all levels of the 1973 revision, great care was taken to insure that each item had both a statistically significant biserial correlation with the total test score of which the item was a part and a progressively higher percent of pupils passing the item from one grade to the next in the level for which it was designed. Items with biserials of less than .30 were eliminated. Typical correlations ran in the mid to upper .40's with some running as high as the .50's and .60's.

The item types are described in the manual as being tests of "word knowledge, verbal analogies, verbal classification, sentence completion, numerical problem solving, number series, pictorial analogies, and pictorial classification," and it is remarked that such items

have stood well the test of time and a large proportion of them survived item analysis in the 70's. . . . By providing an opportunity to solve problems involving these symbol systems, the tests are avowedly measures of general abstract reasoning ability—the type of ability involved in the successful solution of most school-oriented tasks. . . . Such concepts as "social," "practical," or "mechanical" intelligence are not tapped by the Henmon-Nelson tests.

The general purpose and philosophy of the tests are much the same as before:

Like their predecessors, [these tests are] designed to measure those aspects of mental ability which are important for success in academic work and in similar endeavors outside the classroom. Scores may be interpreted as measures of general mental ability, or academic aptitude, but not as indices of school achievement specifically taught in a common curriculum. Although knowledge is required to answer the items, they were selected to reflect ability to profit from exposure to learning in a variety of situations. High performance on the test requires the efficient utilization of verbal and numerical symbols and the ability to acquire and retain information in common symbol form for use at later times in the solution of verbal, quantitative and abstract reasoning problems.

A test of general mental ability . . . is built upon the premise that the separable factors contributing to this ability are themselves substantially inter-correlated. Long-standing research has established that varying degrees of competence in the types of mental operations measured by the Henmon-Nelson tests are in fact significantly related to success in school work. As with any test, however, the results from a single administration should be interpreted as representing the individual's *current level of developed ability* and, as such, are more useful for short than long range prediction. Repeated testings at intervals of two or three years are recommended to establish more stable patterns of performance and bring predictive information up to date.

In attempting to persuade the user that the tests are reliable measures, the manual gives a series of stepped-up odd-even reliability coefficients, generally in the .80s and low .90s, but cautions that this procedure of determining reliability

leaves something to be desired for tests in which speed may be a factor in determining an individual's score. Because there is a definite time limit, the Henmon-Nelson tests do have, for some pupils, an element of speededness.

Thus, the test reviser made no attempt to obtain reliabilities of speeded tests by equivalent-form methods, as he might easily have done, or to take advantage of various methods of determining these reliabilities by statistical procedures suggested by Gulliksen (1950) and others (Guttman, 1955).

The manual proposes to establish the validity of the test by citing correlations between test scores (or IQs derived from the test) and scores on standardized achievement tests. A table (Table 10 in the 1973 manual) shows such correlations ranging from a high of .86 with the Reading subtest of the Iowa Test of Basic Skills, down to .60 with the Arithmetic score on the same test. It is not stated whether these correlations are based on data obtained from testings done at closely similar times (giving concurrent validity) or at widely spaced times, the Henmon-Nelson test being given earlier (giving predictive validity). In any case, the rather high correlation coefficients tend to suggest that the Henmon-Nelson test, with its large complement of printed verbal materials, measures the cumulative effect of school achievement—regardless of how well it measures "general abstract reasoning ability."

Because the 1973 revision of the Henmon-Nelson test appeared too late for treatment in the *Seventh Mental Measurements Yearbook* (Buros, 1972), it is necessary to turn back to the *Fifth Mental Measurements Yearbook* (Buros, 1959) to find reviews of a very similar previous revision (Lamke & Nelson, 1957-1958). Here, we find Leona Tyler (widely known for textbooks on testing and individual differences) praising the revision as exemplifying "the application of the newest technology of mental measurement to one of our oldest testing problems." Similarly, D. Welty Lefever,

a professor of education at the University of Southern California, seems highly favorable to the revision. Both Lefever and Tyler, however, are concerned that the test produces only a single score. As Lefever puts it,

Will a relatively short test (90 items in 30 minutes) carrying a heavy verbal emphasis and yielding a single score serve the purposes for which an intelligence test is being administered? . . . According to the authors, research findings support the position that "factor" tests or multiscore tests have not predicted scholastic success any better than tests producing a single global score. Subtests in multiscore batteries which apparently have little to do with success in a given subject matter field often yield higher correlations with grades in that field than do the subtests which, because of the nature of their content, ought to be the best predictors. . . . If a single predictor of school success is needed which can be given in less than a class period and scored with a minimum of time and effort, serious consideration should be given the Henmon-Nelson test. On the other hand, if guidance involving some differentiation among aptitudes is required and if a profile showing the strengths and weaknesses of each counselee is desired, a more complex test battery will be more appropriate. (Buros, 1959, pp. 470-471)

The Differential Aptitude Tests (DAT). This multiscore aptitude battery, developed by Bennett, Seashore, and Wesman (1947-1969) at the Psychological Corporation, is intended primarily for use in educational and vocational guidance at the secondary-school level. It contains not only tests of several "mental abilities" identified in factor-analytic researches but also several tests that are avowedly measures of certain types of school achievements—English language usage and spelling—on the assumption that these are particularly relevant for guidance decisions. Among test batteries of its type, it is exemplary in the thoroughness with which its standardization and validation have been conducted.

Each of the two equivalent forms of the battery published in 1947 contained eight subtests, taking about four hours to administer: verbal reasoning, numerical ability, abstract reasoning, clerical speed and accuracy, mechanical reasoning, space relations, spelling, and grammar. A revised pair of forms published in 1961-1962 had these same subtests, with the same items, but with a reduction of the number of response options in one of the tests, and elimination of the correction-for-guessing scoring for four of the tests. Thus, the presently available forms are practically identical to those copyrighted in 1947, but the changes in scoring necessitated a revised set of norms. Although the battery is not intended to contain a test of "intelligence," certain combinations of scores will yield high correlations with group intelligence test scores. The battery is accompanied by a manual containing voluminous data concerning test characteristics, norms, and validity coefficients, and by other aids—for example, a casebook—to help high-school counselors learn how to use the test results.

In the manual, the battery is described as a compromise "between the desire to measure 'pure' mental abilities that emerge from factor analysis and the practical necessities continually encountered by personnel and guidance workers through the years." Many of the tests resemble those investigated in factor-analytic studies such as those of Thurstone and others. The manual gives no information, however, about how test items were generated or selected nor does it give any information about whether item-analysis procedures were employed in constructing the tests. Nevertheless, the reliability data given for the tests give assurance that the subscores are satisfactorily reliable for individual guidance purposes.

The DAT battery has received extensive reviews, generally highly laudatory, in Buros's *Mental Measurements Yearbooks* (Buros, 1953, 1959, 1965, 1972). A sample of general comments is as follows: "One of the more valuable tools for sound vocational and educational guidance available today is represented by the Differential Aptitude Tests" (Bechtoldt). "The test items show signs of rare ingenuity. . . . In summary, the DAT have been carefully developed and standardized by competent authors who have done an exceptionally good job in making information about these tests available to the public" (Berdie). "[The DAT] constitutes the best available foundation battery for measuring the chief intellectual abilities and learned skills which one needs to take account of in high school counseling" (Carroll). "Reviewers in the past have justifiably dubbed the DAT as the 'best' available instrument of its kind" (Quereshi, writing in the 1972 *Yearbook*).

Nevertheless, reviewers have pointed to certain deficiencies. Several complained about the lack of information on test-construction procedures. Keats asserted that "it is the responsibility of the test constructor to investigate the question [of test homogeneity] and so to be able to assure users that the test scores have an unambiguous meaning—at least in this sense." "The research," he went on (writing in the 1965 *Yearbook*), "is certainly not beyond present methods of analysis, either theoretically or computationally." Carroll, writing in the 1959 *Yearbook*, complained about the out-of-date standards of English usage that seem to have been observed in the "grammar" test, and Keats, in the 1965 *Yearbook*, pointed out that the 1961–1962 revision took no cognizance of this complaint. Considerable discussion centered on the "differential aptitude" aspects of the battery. As Humphreys noted,

For efficiency in differential prediction, it is desirable to have a battery with low intercorrelations. While the intercorrelations among the tests in the present battery are low enough to make the battery useful, they are not as low as possible. Purer tests than several of those used here are available. (Buros, 1953, p. 681)

Frederiksen (in Buros, 1959, p. 675) wrote: "The question might be

legitimately raised as to now 'differential' the tests are, especially for predicting academic criteria." Schutz (in Buros, 1965, p. 1006) concluded: "Thus, despite the extensive predictive validity coefficients for the separate DAT subtests, the *differential* validity of the tests in predicting various criteria is still without substantiation."

The redundancy of the tests and the inefficiency of using separate subscores have also been criticized. As Carroll noted in the 1959 *Yearbook*:

The authors make much of the fact that the DAT battery is a group of tests, each of which is valid and useful in itself; they appear to advise against the combination of weighted scores. In the notable controversy about clinical versus "statistical" prediction they are on the side of "clinical" prediction. Paradoxically, this position prevents them from displaying some of the undoubted powers of the test in affording statistical prediction from weighted combinations of scores. (Buros, 1959, p. 673)

But Keats points out that the latest manual, published in 1966, contains little information on the possibilities of using multiple correlations and weighted combinations of scores. On the basis of a factor analysis, Quereshi concluded that "three factors generally account for about 90 per cent of the reliable variance, indicating that the battery probably encompasses three factors, but definitely no more than four." He asserted, therefore, that "the DAT, despite its technical superiority over its competitors, represents a substantial degree of inessential duplication of time, effort, and expense." Quereshi also pointed out that the extensive norms, based on more than 50,000 cases, were still not as representative as they might have been. He summarized his feelings in the following words:

It [the DAT battery] has not so far lived up to the promise of a differential instrument, but no other instrument for the given age range has yet succeeded in this respect. However, the DAT, if certain steps are taken, has better chances of attaining an acceptable level of differential efficiency than any other comparative [sic] battery. (Buros, 1972, p. 1052)

The College Board Scholastic Aptitude Test (SAT). The work of C. C. Brigham in developing for the College Entrance Examination Board a test of "scholastic aptitude" to be used for general prediction of success in college and university studies has already been mentioned. Various putatively equivalent forms of this test, known widely to college applicants and their parents as the "SAT," are constructed periodically at Educational Testing Service in Princeton, New Jersey, and administered to more than a million college applicants every year at numerous testing centers throughout the country and even abroad. It represents, or at least potentially represents, the highest standard of test construction technology now available.

The SAT is a "secure" test, that is, the forms (or even the items, except for a few that have been released) are not open to public inspection, and

copies of the test that are actually used are carefully guarded and accounted for. A publication edited by Angoff (1971a) constitutes, in effect, a "manual" for the test and presents voluminous details about the test itself and the procedures by which test items are developed and selected for inclusion in successive forms, successive test components are equated to one another, and norms are established. It also contains much information (even though it is only a sample of the information actually available) about the predictive validity of the SAT and the various factors that affect scores (coaching, practice and growth, fatigue and anxiety, etc.). (It also contains information pertinent to the "Achievement Tests" that are offered by the College Board, but we will leave aside consideration of these tests here.)

For a general description of the test, we could do little better than to offer excerpts from P. H. DuBois's review in the *Seventh Mental Measurements Yearbook* (Buros, 1972):

Like many developments in measurement, the SAT is a direct descendant of the Army Alpha. The original plan was formulated by Robert M. Yerkes, Henry T. Moore, and Carl C. Brigham. Later Brigham became responsible for its development. A conscious effort was made to develop an instrument that would measure neither school achievement nor general mental alertness. With the passage of time, resemblance to its prototype lessened: subtests became fewer; speededness was reduced; and emphasis was placed upon two relatively homogeneous item types—verbal and mathematical. Currently the two sub-scores are referred to as SAT-V and SAT-M.

Even today, nearly three decades after Brigham's untimely death, the SAT continues a number of features which he initiated: (a) an attempt to measure "aptitude for college studies" rather than intelligence; (b) equating of numerous forms so that the predictive scores are stable irrespective of the time and place the student takes the test; (c) extensive and systematic use of item analysis; and (d) tryout of all new material before items are used operationally.

Items with known statistical characteristics are organized in sections requiring 30 or 45 minutes each for a total testing time of three hours. Within each block of items, arrangement is in terms of increasing difficulty, with "the mean difficulty of each block . . . equal to that of the test as a whole."

Test content is selected to avoid as far as possible any inequities to any subclass of the intended population. The SAT-M requires as background only the mathematics taught in grades 1-9, while the SAT-V is related to social, political, scientific, artistic, philosophical, and literary areas. The earlier literary focus of the verbal items has been abandoned for a broader orientation, while the mathematical material now depends less on formal knowledge and more on "logical reasoning and . . . the perception of mathematical relationships."

As is well known, SAT scores are reported with a mean 500 and standard deviation 100. It is not so well known that the standardization group consists of 10,654 students who took the test in April, 1941. Through somewhat involved methods, the various forms are equated accurately to one another and to the form yielding the initial norms.

Ask for any conventional statistic about the SAT and one can be practically certain that it will be available: reliabilities, validity coefficients, and item data. . . . The internal consistency reliability estimates for 12 recent forms cluster

closely around .91 and .90 for the verbal and mathematical scores, respectively; the parallel-form reliabilities average two points lower. In the 1950's the average correlation between the two parts was in the middle .50's; more recently, it has been in the high .60's, possibly reflecting a decreasing emphasis on specific knowledge in the two areas. [It may also reflect an increasing verbal load in the mathematical part, however.]

From the beginning, the SAT has been found to have reasonably good validities for predicting college achievement. Also it has been found consistently that the SAT increases the validity of the high school average or rank. Studies made in 1927 showed a median validity of the school record of .52, a median validity of the SAT of .34, and of the combination of both, .55. . . . Currently, typical validities are .39 for SAT-V, .33 for SAT-M, .55 for the high school record, with a multiple correlation of the order of .62. The picture today is not greatly changed from that in 1927 except that both SAT validities and school grade validities have increased, and the increment of the SAT over the high school record is a little greater than it was. (pp. 646-647)

A few notable details are absent from this account. For example, the SAT test is completely "objective," all items being of the multiple-choice type with five options per item. The exact composition of the sections is not quite as homogeneous as DuBois implies. A typical verbal section, for example, will contain antonym items, sentence-completion items, analogies items, and reading-comprehension items. From a factor-analytic viewpoint, such a section would tap both the "V" (verbal knowledge) factor and a verbal reasoning factor. The mathematical sections consist of two "formally distinct item types: general mathematics items common in form to many other tests, and data sufficiency items" (Angoff, 1971a, p. 17). In one factor-analytic study of the mathematical section, 6 mathematics factors and 10 verbal factors were tentatively identified (Pruzek & Coffman, 1966). The matter of the validity of the SAT in predicting college success is, of course, much more complex than could be conveyed in DuBois's account. One should emphasize that validity coefficients vary widely depending on various circumstances—the type of college, the selectivity of the college, and so forth.

Reviewers of this test in Buros's *Mental Measurement Yearbooks* generally approach it with what almost appears to be reverence and awe. DuBois, whose review (Buros, 1972) has already been extensively quoted, finds little if any fault with the test, and defends it:

The SAT is not without its detractors, some of whom regard it as a "tool of the academic establishment." Of course it is! But it is a good tool, perhaps the best that can be devised with present psychometric technology. No attempt is made to adjust SAT scores for sex, socio-economic status, race, or educational background. In this reviewer's opinion, such attempts would be difficult to justify. Adding substantially to the prediction of academic achievement as contrasted with the use of high school grades alone, the SAT has repeatedly proved its usefulness to colleges. It helps to pick those who fit in best in the program of a particular institution, and, of course, it can be supplemented with any other information that admissions officials may wish to use.

The question has arisen of whether or not it fits the needs of the students as well as those of the colleges and universities. No good end would be served by blunting this instrument. Certainly there are many aptitudes other than verbal and mathematical that some colleges might like to know about, such as musical, spatial, mechanical, and clerical abilities. College entrance examining could be broadened so that a profile of testable characteristics adequate for educational and vocational counseling would become available. . . . Obviously the CEEB program could be broadened further. But even in a broadened program the present SAT measures would be important. (p. 648)

In the same volume of the *Yearbooks*, Wimburn Wallace concludes his review with a comparison of the SAT with its nearest competitor, the American College Testing Program—a comparison that, he says, “favors the SAT overall. . . . Although validities against college achievement criteria are not strikingly different, the content mix, psychometric quality, and pertinent research efforts appear to be superior in the SAT.” His only significant complaint would seem to be that “the content and format of the test have been essentially the same for a very long time.” “The investigation of, or experimentation with, various and new test types for the SAT has been sporadic. . . . There should be sustained, regular effort toward new test development,” he avers. Actually, Angoff’s compilation (1971a, 156ff.) reports numerous studies designed to find new and valid content for the SAT; the fact was, however, that little of use could be found. Various recommendations about possible new types of test content were made by a Commission on Tests appointed by the College Board in 1967 to “undertake a thorough and critical review of the Board’s testing function in American education and to consider possibilities for fundamental changes in the present tests and their use in schools, colleges, and universities” (CEEb, 1970, p. xiii), but thus far these recommendations have not been followed up, possibly because they were seen to be unfeasible.

In earlier *Yearbooks*, we find little criticism of the SAT. In the 1965 issue, Wayne Zimmerman raises the question of why it should take as long as a half day to obtain accurate measurements of just two abilities. In the 1953 issue, Frederick B. Davis (author of well-known reading-comprehension tests) complains that the items

lack originality and interest. The passages on which the reading items are based seem dull and, in general, poorly written. The difficulty of the items based on them often seems to derive from the muddiness of the passages rather than from the necessity for close reasoning or sensitiveness to implications on the part of the examinee. . . . As in the case of most tests, some faulty items can be found. . . . The validity data indicate that this test is about as useful for predicting academic success as a test of its limited scope can be. Tests that have been commercially available for twenty years are about equally useful. Any unique merit in using the *Scholastic Aptitude Test* lies in the fact that one can be reasonably sure that the examinees will not have had access to it beforehand and that it be administered and scored with scrupulous care and accuracy. (pp. 383–384)

It is surely a matter of opinion whether the current forms of the SAT show any improvement over those criticized by Davis in 1953.

2.9 *The Use and Misuse of Intelligence Tests in Schools*

By the opening of what we are calling the "modern" period of the mental test movement, the practice of using intelligence tests in schools, particularly the primary schools, for general diagnosis of pupils' presumed learning capabilities had become well established. Many schools used scores on such tests, along with records of actual school achievement, as a basis for "ability grouping," that is, placement of pupils into groups of "fast," "average," and "slow" learners, apparently with the conviction that pupils' needs could be better served in this way. Bright pupils would presumably be given instruction that allowed them to progress faster through the assigned curriculum units, with appropriate "enrichment" through the introduction of additional material. At the high-school level, such pupils would be more likely to be assigned to college-preparatory courses in mathematics, science, social studies, language, and the humanities. The less able students were given instruction with simpler content and at a slower pace; at the high-school level they were encouraged to take courses in business, shop, and other vocational subjects. The rationale for these procedures was the notion that mental ability scores were indicators of learning capacities that were at least to a large extent permanently fixed in students; the evidence cited as supporting this notion consisted of the many substantial correlation coefficients found between mental ability test scores and school grades. If a student made a low score on a mental ability test, it could be rather confidently predicted that he would do poorly in the more difficult and advanced aspects of school work. Therefore, it was reasoned, the student would be better off, in his general adjustment to school and in his school progress, if he (or she) were assigned to relatively easy tasks and easy subject matter.

Research on the efficacy and usefulness of ability grouping had up to 1935 yielded no clear conclusions and the findings continued to have a negative tone even throughout the modern period (see reviews by Svensson, 1962; Findley & Bryan, 1971). Most of the research on ability grouping, however, pertained to differentiated streams of learners within a narrowly defined curriculum area, for example, groups of fast and slow learners in a particular social studies course. Findings regarded as negative consisted of observations that a slow group made no greater progress as a segregated group than they would have as a nonsegregated group taught along with the fast group. It was rare that researchers in ability grouping had an opportunity to study situations in which instruction was appropriately differentiated for groups of varying abilities.

This research on ability grouping was probably irrelevant to the question of whether students should be guided into different high-school

curricula—academic curricula for the intellectually superior, and nonacademic, “general” curricula for those regarded as having mediocre intellectual abilities. Because of tradition and the weight of school administrators’ experience in student assignments, it was practically impossible to conduct an adequately designed experiment. Such an experiment would have required something like random assignment of representative groups of students to courses of different types of content and difficulty levels.

Actually, if such an experiment had been done, it is highly likely that the usefulness of ability grouping would appear to have been justified to a much greater degree than it was by the usual study with a single subject-matter. That is, we may confidently predict (and school administrators would also have predicted) that most low test-scorers would fail in the more difficult courses, and most high test-scorers would be bored and unchallenged in the less difficult courses. In the 1940s and perhaps in the 1950s, such an experiment (if it could have been conducted at all, in view of practical and ethical considerations) would have been regarded as a supreme vindication of the usefulness of mental test scores as a basis for student guidance and assignment. In fact, as late as 1959, James B. Conant (1959) suggested that only about 15% to 20% of the high-school student population possessed the talents fitting them for college-preparatory courses and eventual entrance to college. In Conant’s own terms of reference, he may have been approximately correct. The difficulty comes in defining the proper terms of reference. That is, if by “possessing talents” it is meant that students have acquired *developed* general abilities, knowledges, and skills needed for success in academic work, one may agree with Conant, but if it was also implied that no other students in the age cohort could have developed those same abilities if they had been exposed to suitable learning and educational experiences from their earliest years, there is a reasonable doubt about the soundness of Conant’s assertion. At least, we shall have to say that the evidence for thinking that only some small proportion of the population has an inherent potential *at birth* for developing “academic talent” is totally inadequate. It is beyond the scope of this chapter, however, to consider this problem in any detail.

It is of interest and relevance, however, to note the impact of research into mental abilities on the British educational system in the early days of the modern period, and the subsequent course of events. The post World War I years in Britain were marked by an often bitter struggle between those who desired to maintain the traditionally highly selective school system (and defend the “public school” system represented by Eton, Winchester, and the rest) and those who sought to reform the system by raising the school-leaving age and making secondary education widely and freely available to children of the working classes (Simon, 1974). There were problems of both economics and ideology. The Education Act

of 1902 had already laid the foundation for a system of publicly maintained secondary schools, but the number of places allotted to non-fee-paying students continued to be very small. While in the early years of the century selection for free places in secondary schools was based mainly on teachers' reports, from about 1924 on, mental tests played an increasingly important role. As Simon (1974) notes,

From the outset tests were integrally related to the selection of a few children from among many at the age of 11, on the grounds of capacity to profit from the academic secondary course. Accordingly the relevant curriculum was taken as given and as the yardstick for diagnosing "intelligence"; the children who could "take it" were intelligent, those who could not lacked ability. There could hardly be a better recipe for crystallising school practice at the secondary stage, and the influence stretched down to crystallise primary schools into a selective pattern. (pp. 239-240)

Intelligence testing in school selection was sold to the public as a scientifically justified procedure (Simon, 1974):

In a popular exposition of the new "science" of mental testing in 1933, to a radio audience, Cyril Burt [psychological consultant to the London County Council] outlined the conclusions arrived at about the functioning of the human mind. It was a simple, clear, straight-forward statement putting no doubts, inviting no argument. "By intelligence the psychologist understands inborn, all-round intellectual ability. It is inherited, or at least innate, not due to teaching or training; it is intellectual, not emotional or moral, and remains uninfluenced by industry or zeal; it is general, not specific, i.e., it is not limited to any particular kind of work, but enters into all we do or say or think. Of all our mental qualities, it is the most far-reaching." The exposition closed with an affirmation with ominous overtones for the educational system: "Fortunately it can be measured with accuracy and ease."

Even within the small world of mental testing, or psychometry, there were differences on this question which might have been cited; notably disagreement with the claim that the *ad hoc* tests, produced to meet urgent administrative needs, measured an innate and unchangeable power of the human mind. One of the most distinguished investigators in the field, Godfrey Thomson, held that there were no grounds to say that the "intelligence" measured by tests was a given quantity transmitted by inheritance; in his view environmental factors accounted for as much as 50 per cent of the variance in test scores. But, once again, so far as the school system was concerned the categorical Burt interpretation was both relevant and requisite, provisos an unnecessary complication. According to this, it was only necessary to take a child's I.Q., or intelligence quotient, when he entered school, allocate him to the relevant stream or differentiated course, and see him through at the given level until he left. This was to plan education in accordance with the child's needs, it was asserted, since it was now known that the I.Q. accurately represented a quota of innate "intelligence" which would not, indeed could not, change. (pp. 241-242; in a footnote, p. 242, Simon points out that Burt expressed similar views in a broadcast talk in 1950)

Since even by the late 1930s the publicly maintained secondary-school system continued to stress an academic curriculum, increasing pressure built up in the general public, particularly in the Labour Party, to estab-

lish a system that would better serve the needs of the great bulk of the students from middle and working classes. To be sure, there were various kinds of vocational and technical schools available, but progressives were of the opinion that the existing secondary schools should have a more diversified and widely appearing curriculum, with a more democratically composited student body. The upshot was the establishment, by the Board of Education, of a Committee on Curriculum and Examinations in Secondary Schools. It was perhaps characteristic of the educational politics of the time that the chairman was a former headmaster of a "public" school, Sir Cyril Norwood, who could not have been expected to be especially sympathetic to progressive ideas. The committee's report was released in 1943. Simon (1974) terms it "the most discreditable of official reports" and remarks that "directed to justifying the Board doctrine of securing a hierarchy of maintained schools in traditional terms, it justified the unjustifiable in the richest of rhetoric" (p. 283).

The report made little mention of intelligence tests ("the major prop of the Board policy on the scientific front," Simon notes), but from the evidence and witnesses listed as being considered or heard by the Committee, it is plain that its conclusions were strongly influenced by then-current factorial theories of intelligence. In a section concerned with "Variety of Capacity," the report (Norwood, 1943) asserts:

The evolution of education has in fact thrown up certain groups, each of which can and must be treated in a way appropriate to itself. Whether such groupings are distinct on strictly psychological grounds, whether they represent types of mind, whether the differences are differences in kind or in degree, these are questions which it is not necessary to pursue. (p. 2)

It goes on to describe three "types" of students: (1) "the pupil who is interested in learning for its own sake, who can grasp an argument or follow a piece of connected reasoning"; (2) "the pupil whose interests and abilities lie markedly in the field of applied science or applied art"; and (3) "the pupil [who] deals more easily with concrete things than with ideas" (Norwood, 1943, pp. 2-3, *passim*). On this basis the report recommended that a tripartite school system be maintained: a "grammar school" for pupils in category 1, a "technical school" for pupils in category 2, and a "modern" school for pupils in category 3. Evidently this tripartite school system was proposed to accommodate the notion of a "general" factor of intelligence along with the notion of a "practical" factor as identified by Alexander (1935) and others (probably similar to what would be called a "Spatial" factor today): Grammar schools would be designed for those high in *g* and verbal skills; technical schools would be appropriate for those high in "practical" and "spatial" factors; modern schools would be suitable for pupils with lesser abilities in *g* and the "practical" factor.

It would appear that many local authorities in Britain actually went ahead with implementing the recommendations of the Norwood report,

although in recent years the trend in Britain has been toward comprehensive secondary schools somewhat similar to those in the United States, Sweden, and other countries. It is not our purpose here to review later developments; we merely note this rather remarkable instance in which research on mental abilities seemed to have had signal influence on the formulation of educational policy, however misguidedly.

In defense of British psychometrists, however, we may note that (in Simon's words) they, "however firmly (if erroneously) they claimed to measure levels of intelligence accurately, were affronted at the crudities of the Norwood typology" (Simon, 1974, p. 327). Burt (1943) attacked the Norwood report vigorously, and numerous subsequent researches had the effect of weakening the validity of Norwood's claims. For example, Lambert (1949) was unable to find clear separations between the presumed types and suggested that "for the whole distribution a bilateral school in which alternative curricula differing slightly in bias would be provided would appear to be more suitable than separate grammar and technical schools." She also found that interests were not fixed at age 11 and that consequently the decision as to what type of school a child might be suited for might be premature at that age.

Simon (1974) comments on the regressiveness of this whole period in Great Britain in the following words:

In sum, throughout the late 1940s and the 1950s, when so much might have been accomplished, the development of secondary education for all was restricted and distorted by the dead hand of a doctrine brought to a point during the depressed 1930s—the doctrine that secondary education is not for all, that only the selected qualify for a complete form of it. . . . In the end it was the breakdown of the supportive ideology of "intelligence" testing, through sheer over-use of tests and exposure of all their inherent weaknesses, an upsurge in the secondary modern schools comparable to earlier outbreaks in the 1890s and 1920s, and growing parental revolt that broke open the closed circle. (pp. 331–332)

This was not to say that "intelligence" tests were completely eliminated, nor that selective grammar schools were completely eliminated. There are still today, in Britain, secondary schools for which one of the criteria for selection is passing of a verbal reasoning test, along with achievement tests in English and mathematics. On the basis of research in one such school, Richardson (1956) asserted that "an intelligence test seems to be necessary," since such a test had significantly higher validity against school grades than any other single test. If one assumes that the test measured some of the generalized developed abilities prerequisite for success in a selective school, its use may have been justified as a practical matter, but it would be wrong to assume that pupils selected by such a test were thereby shown to be innately different from pupils failing the test; it could be that the latter had simply not had the early advantages enjoyed by those passing the test. Furthermore, recent evidence in Britain (e.g., Christie &

Griffin, 1970) suggests, again, that there is no particular advantage in setting up selective schools as opposed to nonselective, comprehensive schools.

In the United States, of course, public secondary schools have been of a comprehensive type for decades, and all vestiges of selection for such schools have long since disappeared. Mental ability tests are used in secondary schools mainly for guidance purposes, if at all—for example, in helping students decide upon choice of courses and curricula.

Test publishers offering group intelligence tests for use in elementary schools have noted a marked decline in sales, since many school systems have followed the lead of the New York City schools in abandoning the use of such tests altogether (Loretan, 1965). Many reasons have been given for the abandonment of group intelligence tests. For one thing, educators feel that teachers are liable to be unduly influenced by the results; that is, test scores are too likely to be used to "type" the child as a slow learner and thereby to inhibit proper instruction. Second, it is felt that group intelligence tests are not sufficiently accurate in view of their speededness and in view of the conditions under which they are usually given. Third, it is thought that the tests yield little information that cannot be gained from well-designed achievement tests in various school subjects. Fourth, it has been recognized, increasingly, that group intelligence tests of the conventional sort tend to measure the kinds of verbal and reasoning skills that, on the average, are more likely to be acquired by middle-class, majority-group children than by children from various types of "disadvantaged" backgrounds; thus, it is said that the tests are "biased" against these latter groups of children, in the sense that these children have had meager opportunities to learn the verbal and reasoning skills that are measured by the tests. In the past, there had been flagrant misuses of intelligence tests, for example, in using English-language intelligence tests to classify children who had little or no knowledge of English.

Granting all these objections to the use of group intelligence tests in elementary schools, one may nevertheless feel that some losses have been entailed in their abandonment. It is not completely certain, for example, that intelligence tests measure nothing beyond what can be measured by achievement tests. The fact that scores on intelligence tests and on achievement tests are generally highly correlated does not necessarily mean that they are measuring "the same thing"; it may only signify that whatever is measured by intelligence tests *tends* to be learned to the same or a similar degree as the skills and knowledges that are specific to a particular subject matter. If group intelligence tests measure certain skills (e.g., general vocabulary and language comprehension, general reasoning power, ability to manipulate symbol systems, etc.) that are not directly taught in schools, an intelligence test—wisely used and interpreted—may yield information that will not become apparent through teacher observa-

tions, achievement tests, and the like. Considering the fact that group intelligence tests have changed little in the last thirty or forty years and have been inadequately reflective of newer developments in test theory, the factor-analytic investigation of human ability, and the study of cognitive growth, it is conceivable that mental tests could now be developed that would have some real value in the process of education.

Mental ability tests in college admissions. Successive generations of college-bound students have been, many would say, all too familiar with the role of "scholastic aptitude" tests, either those of the College Board, those of the American College Testing Program, or those of still other agencies, in college admissions. The evidence assembled by Angoff (1971a) shows moderate to substantial validity of the College Board's Scholastic Aptitude Test (SAT), as given prior to college entrance, in predicting academic success in selective colleges (in terms of grades, retention, and other criteria). If there is any test upon which all feasible applications of test theory have been made, it must be the SAT. It is agreed among test experts that even with all the refinements that test theory, factor analysis, and other psychometric developments can offer, tests of the SAT type can never attain a validity beyond a certain point—represented by a correlation coefficient of about .6 after "correction" for test and criterion unreliabilities and for restriction of range. Much more seems to be involved in college "success" than whatever is measured by the SAT; the SAT is not claimed to be a measure of motivation, personality, creativity, or other traits that might be relevant. The College Entrance Examination Board, and the Educational Testing Service (which prepares and administers the test), has repeatedly cautioned college admissions officers against the exclusive use of test scores in the admissions process. Even when test scores are used judiciously, however, it cannot be denied that they play a considerable role, although the actual extent of this role is difficult to measure or state precisely because SAT scores tend to be correlated with other types of information. Thus, it is likely that the rank order of applicants determined by a selection board *with* the use of SAT scores would be substantially the same as the rank order determined *without* the availability of SAT scores. Nevertheless, it seems that the primary case for using SAT scores rests on the fact that they contribute a significant amount of information over and above the high-school record (Fincher, 1974).

This is not the place to consider the large and controversial literature on the pros and cons of using scholastic aptitude tests in college or university admissions, either at the undergraduate or graduate level. Nor can we comment on the experience of some institutions, like Bowdoin College, that have abandoned the use of such tests in the admissions process. We cannot go into the knotty problems of restriction of range, criterion contamination, cultural bias, the effects of coaching, and other

factors affecting the validity of predictive tests. Many of these matters are treated quite thoroughly in the technical manual for the SAT compiled by Angoff (1971a). It may be pointed out here, however, that the scientific examination of all these problems is a challenge that draws upon the full resources of statistical methodologies and theories of measurement that have been developed over the years.

From the standpoint of the theory of mental abilities, furthermore, we may offer several comments:

1. To the extent that tests of the SAT type are valid in predicting college or graduate-school success, it is undoubtedly because they provide a good indication of the extent to which applicants have *at the time of testing* developed or acquired, and can exhibit through their performance on a test, certain general intellectual skills in handling verbal, quantitative, and symbolic information that are contributory or even necessary to high-level success in academic studies, and, in addition, to the more intellectual types of pursuits in postcollege years.

2. A low score on such a test is no guarantee that the individual cannot acquire, during a subsequent period, the skills and abilities that are tested, but it is an indication, at least, of the probability that the individual could acquire these skills and abilities, if at all, only with much expense of time and effort and with careful instruction. This statement is an inference from at least two types of evidence: First, it has been repeatedly shown that coaching in short courses aimed at increasing SAT scores has meager effects, if any; second, the increases in mean SAT scores that occur from the third to the fourth year of the secondary school are very small. Furthermore, I can find little evidence that special compensatory or remedial courses for low-scoring students conducted for a period of a year or two in colleges have any major effects in increasing the level of individuals' academic abilities, although they may indeed increase individuals' abilities to cope with the demands of academic studies (Gordon & Wilkerson, 1966; Rayburn & Hayes, 1975).

3. Nevertheless, there is actually relatively little knowledge about the extent to which the aptitudes and abilities measured by scholastic aptitude tests can be improved by special training and guidance. The fact that previous efforts to improve them have been generally disappointing does not demonstrate that no such efforts can be successful.

3. REVIEW AND ASSESSMENT: THE CURRENT SCENE

In the preceding sections we have been concerned with presenting the rather checkered history of theory and practice in group mental ability testing. Obviously, an enormous amount of effort has been devoted to the development, analysis, and application of measures of mental ability. But has all this effort been properly oriented to achieve socially useful results?

Has it been based on a correct set of assumptions about the nature of the human organism—its biological substrate and its capacities for learning and adaptation? Has the testing movement been founded on acceptable presuppositions about the nature and goals of education in relation to the goals of human society?

These are very difficult questions. In part this is because they involve individual and societal values which appear to have changed significantly over the period we have reviewed and about which there can be much discussion and debate among different sectors of the public. In part the difficulty of these questions is due to the incompleteness of our knowledge about individual behavior and the workings of education and society. A complete discussion of them would take us far beyond the scope of this chapter, which is limited to considering what contributions the theory and technology of testing might make to the resolution of current issues and problems. Nevertheless, they are questions that must be faced in thinking about the future of the mental testing movement and the kind of public support that it may or may not deserve.

In arriving at some tentative conclusions and recommendations, it may help to consider some of the issues raised by the controversies that have surrounded the testing movement.

3.1 Public Controversies About Mental Testing

Public controversy about mental testing and its consequences is not new; it goes back at least five decades, as Cronbach's (1975) thoughtful essay recounts. In the 1920s, debate was centered around questions of invidious ethnic comparisons, with overtones of doubt about the supposed immutability ("constancy") of intelligence and the elitist implication that education could not erase or even reduce intellectual inequality. One senses that during the whole period from the 1920s to the period of the 1960s, when debate became extremely public and acrimonious, there was a latent, seldom-expressed concern over the use of intelligence tests in schools to sort children into ability groups, to bar students from the more selective schools and colleges, or, to put it bluntly, to determine young people's careers and eventual fates.

In England, the use of the so-called 11+ examinations for grammar-school selection increasingly disturbed the more liberal-minded sectors of the population, even though the tests were sometimes praised because they did, at least, provide a basis for admitting to grammar schools some "working-class" children who might otherwise have been turned away. Young's (1958) satiric essay on an eventual "meritocracy" of the intellect was in part a response to the great emphasis on intelligence testing in British education.

In the United States, an analogous phenomenon occurred, but at the college level, as many colleges found themselves forced to be more selec-

tive than previously (Hills, 1971, p. 681) and increasingly emphasized scholastic aptitude test results in the selection process. Chauncey and Dobbin's (1963) book on mental testing, addressed to the general public, probably did little to allay concern about testing in education, even though these authors asserted that "intelligence tests . . . measure only the individual's capacity for learning" (p. 21), that they "measure not innate ability but a developed ability in which innate ability and learned behavior are mixed in unknown proportions" (p. 22), and that "scholastic aptitude tests include verbal and numerical content but omit the non-academic tasks . . . [and] have all of their content devoted to skills taught in schools" (p. 32). Public concern about mental ability testing has been at least partly responsible for the abandonment of the use of intelligence tests in many public school systems throughout the country, and the use of ability grouping in schools has recently been put under severe restrictions by court decisions asserting that ability grouping limits educational opportunities. Currently, there is considerable feeling in some quarters that the use of scholastic aptitude tests in college admissions should be sharply curtailed or at least put under certain controls that are felt to be necessary.

While there seems to have been general public acceptance of the need and utility of mental ability and aptitude testing in the armed forces during World War II, criticism of the overuse, misuse, and possible discriminatory effects of such tests in business and industry became frequent in the 1950s and 1960s, leading eventually to legislation and government regulations that had the effect of mandating demonstrated validity of any employment selection tests, including mental ability tests, that might be used (Ash, 1966; Baxter, 1969).

Much of the debate during this period stemmed from the civil rights movement and criticized the possible "cultural bias" of mental tests. It was asserted that the tests were oriented to tapping abilities and achievements valued in white, middle-class education and culture, and that persons in various minority groups, with different cultures, or even different dialects and language backgrounds (Labov, 1969), were put under a great handicap by the tests. At the 1963 meeting of the American Psychological Association, a group of black psychologists presented a manifesto calling for a moratorium on the use of psychological and educational tests in schools with students having disadvantaged backgrounds, presumably until further research could be done to correct defects in tests and testing procedures.

During the 1960s, the Russell Sage Foundation initiated studies of the social effects of testing. As Orville Brim notes in his preface to the first volume (Goslin, 1963) of a series of reports,

the increasing use of tests in the United States constitutes a change in emphasis from traditional bases for the determination of status, such as race, sex,

religion, and order of birth, to a greater reliance on this new criterion, performance on a standardized test of ability. . . . The source of this change can be traced to a number of factors, including the growing concern for education in America and the greater technological complexity of the society which makes it of increased importance that individuals occupy positions for which they are well suited. It is notable that the growth of standardized testing has come about as a direct result of the application of social science techniques to the solution of a problem; that is, how to select the best-qualified individuals for the various educational and occupational positions in society.

At present virtually nothing is known about what effects the testing movement is having on our society, and on the individuals who are directly affected by test results. The social consequences of standardized testing have not been considered systematically by testers or by social scientists, despite the expanding importance of the movement and the number of decisions being made on the basis of test results. (p. 3)

Goslin's volume set the stage for the studies by reviewing evidence available at the outset, discussing the possible favorable and unfavorable effects of mental testing on the individual and on the society. Echoing Young's (1958) forecast of a meritocracy of intellect, he stated: "One conceivable consequence of a greater reliance on tested ability as a criterion for the assignment of educational or occupational status is a more rigid class structure based on ability" (p. 191)—an idea that was to be further expanded by Herrnstein (1973). On the basis of extensive interviews, Brim, Neulinger, and Glass (1965) found that adults' estimates of their intelligence were more frequently based on "success in work" than on intelligence test scores; the tests were seen to measure mostly learned knowledge, and to be "accurate." These findings, at least, were reassuring in some ways, as were also those of Holmen and Docter (1972), who in a study of the "testing industry and its practices" reported that

generally, improvement of the instruments has met with more success than improvement of processes at the user level. . . . The good practices at the test-research and development level are not sufficient at this time to offset testing system inadequacies at the test user level. (p. 171)

Holmen and Docter were thus more concerned about the uses and misuses of testing than about the theoretical foundations of such testing; nevertheless, they present a thoughtful account of criticisms of testing that bear on its theoretical aspects.

Public debate grew to orchestral proportions chiefly as a result of a famous (or, some would say, notorious) article by Jensen (1969), which among other things asserted that (a) compensatory education for disadvantaged groups had "apparently" been a failure, (b) various lines of evidence "make it a not unreasonable hypothesis that genetic factors are strongly implicated in the average Negro-white intelligence difference" (p. 82), and (c) the race differences were evident chiefly in a "Level II," "conceptual ability," and not in a "Level I," "associative ability." In the public outcry against the alleged "racist" connotations of these statements,

the extensive and generally defensible scholarly review that Jensen offered of the theory of intelligence and its measurement was ignored. Jensen (1972, 1973a, 1973b) has, however, shown remarkable persistence in defending his views on the heredity-environment issue that he almost unwittingly brought to public attention. He has been joined by others who lean to genetic explanations, like Eysenck (1973), while the antihereditarian forces have been led by Kamin (1974), an experimental psychologist with no previous involvement in the testing movement, who asserts that "there exist no data which should lead a prudent man to accept the hypothesis that I.Q. test scores are in any degree heritable" (p. 1). The debate at this writing seems to have run its course, although the CBS television program on the "IQ Myth" hosted by Dan Rather (Ravitch, 1975) and the statements made on a PBS NOVA program, sponsored by the American Association for the Advancement of Science, by geneticist Richard Lewontin to the effect that Jensen's conclusions were racist and scientifically invalid have certainly added fuel to it, and the latest "flap" has been the allegation that widely cited twin data assembled by Sir Cyril Burt were fraudulent (Wade, 1976).

Almost paradoxically, the most recent public concern in the area of mental ability testing has been over the decline, noted by Educational Testing Service and the College Board, of the average SAT scores made by college admissions applicants over the past decade. This concern has expressed itself, in my opinion correctly, not in alarm over a decline in the nation's supply of innate intelligence but in distress about declining standards in schools, decreasing motivation and effort on the part of students, and other possible factors in the test score decline. There is as yet no convincing evidence as to the causes of the decline, although a study by Harnischfeger and Wiley (1976) points the finger to curricular changes in schools, among other things.

A comment, written by a younger generation philosopher of science with a Marxist orientation, encapsulates some of the issues that are at the center of attention in present-day debate. Putnam (1973) is discussing aspects of psychology that cannot, he feels, be interpreted in a reductionist framework but, rather, must be examined in terms of social organization and societal beliefs:

As an example, let us take a look at the concept of "intelligence"—a concept in vogue with racist social scientists these days. The concept of intelligence is both an ordinary language concept and a technical concept (under the name "IQ"). But the technical concept has been shaped at every point to conform to the politically loaded uses of the ordinary language concept.

The three main features of the ordinary language notion of intelligence are (1) intelligence is hard or impossible to change. When one ascribes an excellent or poor performance to high or low skill there is no implication that this was not acquired or could not be changed; but when one ascribes the same performance to "high intelligence" or "low intelligence" there is the definite

implication of something innate, something belonging to the very *essence* of the person involved. (2) Intelligence aids one to succeed, where the criterion of "success" is the criterion of *individual* success, success in *competition*. It is built into the notion that only a few people can have a lot of intelligence. (3) Intelligence aids one no matter what the task. Intelligence is thought of as a single ability which may aid one in doing anything from fixing a car or peeling a banana to solving a differential equation.

These three assumptions together amount to a certain social theory: The theory of elitism. The theory says that there are a few "superior" people who have this one mysterious factor—"intelligence"—and who are good at everything, and a lot of slobs who are not much good at anything.

The IQ test was constructed to preserve the elitist features of the concept in the following way. (1) The IQ test was standardized so that IQ scores would not change much with age, thus preserving the illusion of a measure of something unchanging. . . . (2) The IQ test was "validated" by selecting the items so that they would predict "success" in college—[and this] is 100% a statistical artifact of this method of validation. (3) The third feature of the ordinary language concept—that IQ is a *single* factor—was harder to ensure. All of the statistical evidence turned out to be against this hypothesis. In fact it turns out that over a hundred different factors contribute to one's score on IQ tests. So one just takes an average, weighting the factors so that they predict success in school, and *calls* the result "*Intelligence Quotient*." And . . . and behold! One has people with "high IQ" and people with "low IQ," "gifted people" . . . and "dull people." . . . In short, one recovers the full ordinary language use of the concept—but now with the appearance of "scientific objectivity." (pp. 141–142)

Along with the misunderstandings and misinterpretations to be found in this excerpt there are points that need serious discussion. One hopes that Putnam does not mean to imply that *all* social scientists who study "intelligence" are "elitist" or "racist," if indeed definitions of those terms could be agreed upon. One must question his assumption that the validation of intelligence tests is "100% a statistical artifact" as well as his statement that "over a hundred different factors" contribute to one's score on an IQ test. Not even Guilford (1967), whose work Putnam seems to be referring to, would agree with that statement. Also, one finds it hard to square Putnam's idea that the ordinary language concept of intelligence includes whatever it takes to peel a banana with his complaint that intelligence gets defined in terms of scholastic success in college. Nevertheless, central issues raised or implied in Putnam's discussion are the following:

1. The question of whether intelligence is a valid concept, and, if so, whether it corresponds to a "single ability" that generalizes to all kinds of activities, or to a multiplicity of abilities.
2. The possible genetic basis of intelligence, and its modifiability.
3. The relevance of intelligence to scholastic success.

These have in fact been the predominant issues in the history of criticism of the testing movement. There have of course been many others, such as the allegation that the testing "industry" is a sort of cartel

that exerts undue control over people's lives, but discussion of such issues is beyond our present scope.

What lessons are to be learned by specialists in mental measurement from public controversies on their subject matter? Cronbach (1975) has written:

Inquiry is best left unrestricted. But the person publishing or popularizing a study does have responsibility for anticipating what his words will suggest to the rightists and leftists, the exploiters and the *descamisados*. He is not irresponsible when his conclusions sway public decisions; he is irresponsible if his careless writing does so.

Our greatest difficulty is our innocence. To spotlight one question, pleased that social science can answer it, often casts closely related questions into a deeper shadow.

The testers of the 1920s could conceive of no risk or error save that of failure to take the tests seriously. The spokesmen for tests, then and recently, were convinced that they were improving social efficiency, not making choices about social philosophy. Their soberly interpreted research did place test interpretation on a more substantial basis. But they did not study the consequences of testing for the social structure—a sociological problem that psychologists do not readily perceive.

The social scientist is trained to think that he does not know all the answers. The social scientist is not trained to realize that he does not know all the questions. And that is why his social influence is not unfailingly constructive. (p. 13)

Let us be a little more specific. The social scientist has the responsibility not only to conduct his research according to the most advanced methodological canons and in the light of the most advanced knowledge that has accumulated in his field, but also to view his work in the broadest possible context of psychological and social theory. Factor analysts and test theorists have not always been mindful of this latter responsibility. There has been a tendency among them, for example, to hypostatize *ability* as something to be taken for granted, and defined merely by the carrying out of certain measurement and statistical operations. They sometimes fail to remind themselves that the concept of ability must be defined in terms of process rather than structure, and that ability is a manifestation or characterization of behavior that must be described and explicated in terms of its antecedents, not only those of a biological character but also those that stem from environmental experiences. While factor analysts and test theorists need not be specialists in psychological theory, they should be thoroughly grounded in it.

More immediately, what should be the response of measurement specialists to public controversy? Many, of course, will choose to continue doing the things they want to do and are competent to do, and they have a right to do so, in the interests of scientific endeavor. But the interests and competences of others may be such as to allow them to shape their

work to provide knowledge or technology that will be responsive to issues raised in current controversies, along lines that I shall suggest below.

3.2 *New Directions for Mental Test Research*

Possibly in part as a result of public controversies, the interest within psychology in the study of mental abilities is today as strong as it ever has been. New texts and treatises, edited volumes, symposia, and other writings on the subject abound (Brody & Brody, 1976; Cancro, 1971; Green, Ford, & Flamer, 1971; Green, 1974; Husén, 1975; Resnick, 1976; Royce, 1973; Tyler & Wolf, 1974). A new journal that started publication in 1977, *Applied Psychological Measurement*, proves to be largely devoted to issues about mental abilities, and yet another journal, entitled *Intelligence*, has just made its appearance. Psychologists are, or should be, challenged by such statements as one made by the editors of the journal *Cognition* that "there is no theory that unites the higher mental processes, so apparently mental tests are basically uninterpretable," although it is to be noted that these writers are puzzled as to "why there is so much interest today in [research in individual differences] among otherwise serious scientists" (Mehler & Bever, 1973, p. 10). One can be assured, in any case, that competent personnel are in place to perform new research in mental ability testing.

3.3 *Theories of Mental Abilities*

Even the more recent developments in factor analysis that have been reviewed above—Guilford's Structure of Intellect model and Cattell's hierarchical model that specifies "fluid" and "crystallized" types of intelligence along with a traditional general factor—leave much to be desired in the way of a satisfactory theory of mental ability. Guilford's model is logically flawed, and vague in its definitions. Cattell has come much closer to offering a satisfactory psychological theory, with an elaborate array of special terminology and references to presumed psychological tendencies. Nevertheless, I believe, all this work suffers from several major flaws that are not commonly recognized.

First, the "factors" that are identified by the analytic techniques remain as hypostatized end products of the research, rather than being considered as provisional intervening variables to be explained in terms of the dynamics of mental processes, and of antecedent conditions in the individual's constitution and life experiences. There is little *theoretical* interest in any factor, be it "general," "primary," or "group," unless one can offer an account of why the various performances loaded on the factor are correlated or otherwise linked. The finding of a correlation between two performances requires further analysis, for it could arise for any one of a number of reasons. The responses involved might have been learned

together (for some quite arbitrary reason), *or* learning one of the performances might be prerequisite to the learning of the other performance, *or* the performances might indeed both depend upon some trait or capacity (like visual acuity or auditory sensitivity, or "thought power," if such exists).

Second, the data used in practically all the studies consist mainly of scores from tests and measurements that conform to very traditional formats. The tests usually consist of a collection of items that the researcher believes to have some homogeneity, but the actual degree of homogeneity is seldom investigated or even questioned. The scores are obtained by counting the number of items that the examinee has responded to "correctly" within some more or less arbitrary time limit. The responses are obtained at essentially one point of time, there being no effort to assess parameters of learning, memory, problem solving, concept formation, etc., over periods of time. Thus, the kinds of materials that go to make up a factor analysis battery are usually limited in their diversity, even though superficially they may appear quite diverse in content. If a "general" factor is found to underlie the test intercorrelations, it is possibly rather narrow in its generality.

In the case of many studies where general factors are found, it would be hard to dismiss the hypothesis that the general factor simply reflects the fact that all its tests depend on the extent to which examinees have somewhere, sometime, mastered and retained the skills required to perform these tests—language skills, skills of attention and concentration, problem-solving algorithms, facts about the number system, etc. The general factor could then be simply a measure of total learning; it could be attributed to an innate trait only if one could rule out *all* other possible variables that could account for differences: differential exposure to learning opportunities, different styles of child rearing and school instruction, etc. Thus, when Eysenck (1973, p. 480) cites correlations among Thurstone's PMA tests as giving "the firmest support for the existence of a general factor of intellectual ability," he seems to ignore the possibility that the general factor could arise through the fact that people inevitably differ in the richness and variety of the intellectual experiences to which they have been exposed and from which they have profited.

The more general point is that researchers in mental abilities have rarely looked carefully at the measurements they take, tending to regard test scores as givens. This is true of factor analysts, with rare exceptions, and it is also true of test theorists, who seem mainly concerned to put together items, of whatever character and source, into a test that will have certain desirable properties—high reliability, normal score distribution, and so on. Take the matter of speededness in tests: Study after study in the factor-analytic literature, even to the present day, exhibits neglect of the fact, demonstrated by Davidson and Carroll (1945) and Lord (1956),

that scores on tests with restricted time limits can be, and usually are, a function of both speed or rate of work and level of mastery. It is likely that the general factor that is often found in these studies is to some extent a function of individual differences in rate of work that enter into the performance of every test. It is encouraging, however, that papers by Furneaux and by White, reprinted in Eysenck's compilation (1973), have redirected attention to this problem.

What is implied here is that we must make more detailed analysis of tests as complex tasks; I have suggested some ways of doing this (Carroll, 1976). More detailed analysis of tests will not be sufficient, however, unless more research is directed toward finding out how people learn to perform these tasks, and what conditions are favorable to that learning, whether in childhood or later on. Mental ability test scores are normally thought of as predictors of some kind of later learning success; too infrequently are they thought of as dependent variables exhibiting variance that demands explanation. For example, we have few studies in the genre of the one performed by McCullough (1939), who sought to find out how much improvement in certain group intelligence test scores could be produced by a 10-week course in reading. For most of the factors identified by Thurstone (1938a) we have little research evidence about the modifiability of test performances through education, training, or other types of interventions. Marjoribanks (1974) showed that home and school environmental variables predicted much of the variance in the PMA tests.

There are many other things that could be said, but cannot be said here, about ways in which psychometric specialists could reorient their research towards the formulation of a satisfactory theory of mental abilities. The needed technology is highly developed and readily available; what seems to be lacking is a general acceptance and understanding of the task.

3.4 The Heredity vs. Environment Issue

The standard view within the testing profession, and among psychologists generally, has been for a long time that all measured mental abilities are a function of *both* genetic and environmental determinants, the contribution of neither of these being zero. Estimates of the heritability of intelligence range from about 40% to the figure accepted by Jensen (1972) and Eysenck (1973), 80%, unless we accept the figure suggested by Kamin (1974), namely, 0%. The debates between Kamin and Jensen, and their respective supporters, have been so voluminous and technical that following and evaluating them is a task that could require the almost full-time attention of a specialist in genetic psychology. I do not claim to be such a specialist and therefore must rely on reviews and other writings that tend to make me believe that although Jensen's views must be accepted with some caution, Kamin's claim that the heritability of intelli-

gence cannot be shown to be greater than zero is unsupportable. Fulker (1975, p. 519), for example, states that Kamin's account "lacks balanced judgment and presents a travesty of the empirical evidence in the field." Sandra Scarr-Salapatek (1976, p. 99) calls Kamin's book "a disservice both to science and to the advancement of social equality." I am therefore not swayed from accepting the standard view mentioned above, although I find it difficult to accept a heritability estimate as high as 80%.

Before completely dismissing Kamin, however, I should remark that I have always felt that it would be difficult to assess the genetic influences on test scores that are so patently functions of specific learning experiences as typical IQ tests, with their burden of verbal and symbolic material. To that extent I would agree with Kamin that "there are no data sufficient for us to reject the hypothesis that differences in the way in which people answer the questions asked by testers are determined by their palpably different life experiences" (p. 176). But then, so would nearly all psychologists, unless Kamin meant that the differences were determined *only* by life experiences. Kamin continues, however:

That conclusion is silent with respect to another possible question. There may well be genetically determined differences among people in their cognitive and intellectual "capacities." To demonstrate this, psychologists would have to develop test instruments that provide adequate measures of such capacities. They have not as yet done this; they have only developed I.Q. tests. This book has been about—and only about—the heritability of I.Q. test scores. (p. 176)

Here Kamin must be corrected: Psychologists have developed more than IQ tests—they have developed tests and procedures for measuring a great variety of mental processes. But it just so happens that most of the relevant evidence on heritability, that is, in studies of twins and other kinds of kinship, is based on IQ scores, or scholastic aptitude and other tests that are closely similar to standard intelligence tests—for example, data of Loehlin and Nichols (1976) on National Merit Scholarship Qualifying Tests. There are very few relevant data based on tests that depart from the content of typical intelligence tests, such as tests of primary mental abilities (Vandenberg, 1962; Humphreys, 1974). In future studies of heritability, it would be desirable to make more use of tests of differentiable mental abilities, or tests of capacity, learning rate, and the like. Vandenberg (1976) has already claimed to find genetic influences on learning tasks.

Given the difficulty of controlling environments, it will probably be very hard to provide convincing evidence for genetic influences even through twin or kinship studies. Longitudinal studies of twins reared apart are all but impossible. I would attach much more meaning and practical significance to studies designed to determine the actual degree of modifiability of the traits measured by mental tests, and the limits to which abilities could be increased. The technologies of factor analysis and

test theory could make a significant contribution to such studies by providing tests and measures with desired characteristics of independence and homogeneity. If we find that tested abilities can be modified, we would know more about environmental effects and would acquire information of practical value in education. If they cannot, we would be inclined to favor a genetic hypothesis about them; we would at least acquire information useful in shaping educational and social policies.

Another promising line of attack on the heredity-environment issue, as it affects mental ability testing, is through studies of growth, using measures of differentiable mental abilities. Previous studies and analyses (Anderson, 1939; Bloom, 1964; Thorndike, 1966; Jensen, 1973a) have employed global IQ tests, for the most part. Just before his death, Thurstone (1955) completed a study of growth rates in Primary Mental Abilities, concluding that abilities increase at different rates; this kind of work needs to be followed up. It should be noted, however, that the so-called overlap hypothesis explored by Anderson (1939) and Bloom (1964), whereby ability appears to grow by random increments, is not as favorable to an environmentalist position as its advocates may think; it presents the problem of a "starting endowment," among other things.

3.5 *The Problem of Relevance and Validity*

Scores on mental ability tests tend to have at least moderate correlations with school and college academic grades, with the intellectual level of the occupation that the individual eventually follows, and, sometimes, with the level of success in that occupation. Why is this? The standard view in the profession is that the performances on tests are predictive of later performances because they measure characteristics that are also relevant, in some way, to those later performances. In fact, in the development of any aptitude test it is standard procedure to identify, by a "job analysis" of the criterion performance, the kinds of abilities, knowledge, and skill required for successful performance, and then to develop tests to measure those characteristics. The validation of these tests, that is, the demonstration that they correlate significantly with criterion performance, is truly a form of hypothesis testing. There is nothing artifactual in the process, *pace* Putnam (quoted above).

There are, however, several weak links in the system. Tests are never perfect predictors, of course; there are both positive and negative errors of prediction, even though on the whole the predictions are correct. But the system depends crucially on the relative stability and unmodifiability of the aptitudes or other characteristics measured by the predictive instruments. One reason that scholastic aptitude tests are successful predictors is that the verbal and mathematical abilities they measure appear to be relatively stable and resistant to improvement, at least in the short term.

In the history of the mental testing movement, it has always been assumed that better predictive devices could be developed by using more exact and advanced knowledge of the nature of human abilities. This in fact has been, and continues to be, one of the practical motives behind the development of factor analysis, in both its methodological and its substantive aspects. It is also a motive for the improvement of test technology through the theory of tests.

Other than suggesting further work along these lines, we have no really new ideas about how the relevance of mental abilities in criterion performances can be better established. Some would even question the worthwhileness of that task. We may, however, point to several useful and in some ways novel ideas about aptitudes that have come to the fore in recent years. In 1963, I proposed to define aptitude in terms of the relative amount of time that an individual would require to learn a task, subject matter, or skill under optimal instructional conditions; I regarded aptitude, defined in this way, as a function of basic traits and capacities as identified in factor analysis (Carroll, 1963). This idea has never been systematically tested, although I applied it (Carroll, 1974) to demonstrate that long-term school achievement data could be interpreted in terms of rates of progress as predicted by an aptitude measure. Bloom (1976) has modified this notion by characterizing aptitudes as "entry behaviors," and claims to have shown that these entry behaviors can be modified—in the sense that under appropriate instructional conditions students' rates of learning can be made to become more uniform, with the result that individual differences in terminal performance in a task are much reduced. Whether this idea could be applied in modifying aptitudes for school and college success is a question that should be studied extensively, using all the technology of factor analysis, test theory, and experimental design that is available.

3.6 *Other Problems*

Space does not permit an adequate review and discussion of the relevance of mental test technology to a number of other problems, such as that of possible cultural bias in tests used with minority groups, the decline of scholastic aptitude test scores, or the misuse of tests. To a large extent, such problems may be subsumed under the central problems already discussed. For example, the problem of cultural bias should be resolvable by considering it within the general theory of mental abilities—looking into which traits and behaviors are most affected by cultural differences, investigating ways in which those traits and behaviors can be measured so as to minimize biasing effects, or studying ways of improving those traits or minimizing their effects in criterion performances.

3.7 *The Theory-Practice Interface in Perspective: A Concluding Statement*

The notion that individuals differ in intelligence, mental capacity, or mental ability, and that such differences can be measured and studied scientifically, has been one of the important ideas of Western civilization for more than a century. For better or for worse, it has been particularly influential in the shaping of educational policy and in the actual management and conduct of schooling. The study of the development of this notion affords an interesting case in the history of science. Yet, the complexity of the issues and the semantic overloads in the very term "intelligence" make it difficult to frame a discussion that is intended to be as judicious as possible on questions of what intelligence really is, what the relative influences of heredity and of environment on intellectual differences are, and how much mental abilities can be modified by human interventions. In this discussion, I use the term "intelligence" in the broadest and most neutral sense.

When the concept of intelligence as an object of scientific study was introduced by Galton in the late 19th century, it was intimately associated with the notion that intelligence was in large measure an inherited trait. It was natural that Galton's studies of the breeding of plants and animals, and of the inheritance of certain physical characteristics in human beings, led him to believe in the hereditary nature of intelligence. It was probably this belief that caused him to neglect the role of learning and experience in the formation of mental ability.

After the first hesitant and largely unsuccessful attempts of J. McK. Cattell and others to measure intellectual traits through simple sensory and perceptual tasks, the concept of intelligence started to have an influence on school practice with Binet's development of procedures for diagnosing the mental conditions and prospects of slow learners. While Binet did much to define intelligence as the manifestation of cognitive processes, he disavowed interest in the question of genetic origins of intelligence; in fact, he hoped that means could be found for improving the intelligence of mental defectives. Nevertheless, nearly all of those who used Binet's scale or its numerous adaptations were quick to assume that the mental growth rates that it measured were greatly influenced by heredity. Increased scientific respectability was lent to the concept of intelligence as an inherited trait by the statistical studies of the Galton-Pearson-Spearman school, and it is significant that the study of intelligence and its measurement was the principal catalyst for the further development of statistical methods generally useful in the biological and behavioral sciences.

The gradual acceptance of a scientific concept of intelligence or mental

capacity as a stable and probably inherited trait in individuals of all ages led to its use as a basis for mass testing, first in the selection of military personnel in World War I, and then in the schools. As an alternative to the individually administered Binet-type tests, group-administered paper-and-pencil tests were devised; in this process, subtle and largely unnoticed changes came about in the measurement properties and even the theoretical underpinnings of the tests. As methods for validating and standardizing tests became more refined, intelligence tests were eagerly seized upon by educators as a means of identifying promising students for selective schools and of assigning students to different levels of school-work difficulty. Intelligence testing was recognized as a scientific endeavor, and the notion of intelligence entered the public mind as a factor to be considered in the formulation of educational and social policies.

Examination and criticism of the concept of intelligence, either in or outside professional circles in mental testing, were rarely deep. Outside the profession, debate was often based on misunderstandings of the facts and theories offered by test developers, and in any case a general science of human characteristics and behavior had not progressed sufficiently to serve as a basis for sound public discussion of the implications of mental ability testing. Within the profession, argumentation centered around the generality and relative immutability of intellectual differences, but techniques and models for studying these matters were simplistic, and alternatives or supplements to genetic explanations of data were not thoroughly explored. The claims of behaviorist psychologists, such as Watson, that human behavior is infinitely malleable were disbelieved, or patronizingly tolerated as "the *reductio ad absurdum* of a mindless environmentalism gone rampant" (Kamin, 1974, p. 178). Even in the public mind, environmental determinants of individual differences in intelligence were hard to believe in view of the obvious differences that could be observed among children in the same family, with presumably similar upbringing and advantages, and, at the same time, the obvious similarities in the mental capabilities of identical twins. Meanwhile, Brigham's recantation, in 1930, of his earlier hereditarian views and his development of a concept of "scholastic aptitude" distinct from intelligence were not well understood either by the mental testing profession or by educators in general. Tests perceived to be intelligence tests—regardless of the names given them by their creators—continued to be produced in standard molds, periodically revised, and utilized in great quantities. Perforce, nearly all studies of the relevance of mental ability tests for selecting and assigning students in schools, or of the effects of genetic and environmental factors on test performances, used scores on global tests of intelligence or scholastic aptitude as the variable of chief interest.

In the period since the middle 1930s, two lines of work have dominated the scientific study of individual differences: factor analysis and test

theory. Factor analysis has been concerned with the decomposition of mental abilities into "factors," "primary mental abilities," or similar constructs. Test theory has been concerned with various statistical models and operations designed to produce certain desirable measurement properties in tests or test scores. In retrospect, these two types of scientific endeavor have had both productive and counterproductive features. Factor analysis has had at least the potentiality of facilitating the detailed understanding of what could be measured, and test theory has promoted the more accurate measurement of whatever could be measured. On the negative side, however, in factor analysis there has been too much interest in the identification of factors as end products of the research effort, and too little interest in explaining the factors identified in terms of their composition and antecedents. In test theory, there has been overemphasis on desirable measurement properties without corresponding consideration of what exactly was being measured—a pure trait or a mélange of correlated response tendencies. The two specialties drew apart, each failing to take adequate advantage of the achievements of the other. Both specialties developed in such technically complex directions that their findings and theories became all but unintelligible to the practical test user, and even to many persons honestly seeking to construct better and more valid tests.

Despite a few promising signs in recent research efforts, the present scene in intelligence testing is essentially one of stagnation, with much talk but little progress. Buffeted by public criticisms of testing and the shocks of controversy about alleged elitist and even "racist" motives of the testing movement, the profession has all but ceased working toward the development of better and more soundly based theories of intelligence and its uses. The overall decline in the utilization of intelligence tests in schools, the legal constraints on research with human subjects, the general public resistance to testing, for whatever reasons, and the inertia within the profession itself that has resulted from complacency about the sufficiency of presently available methodologies and testing procedures and disquiet about misuses of these procedures have made serious inroads on the ability of the testing profession to move forward. But large sectors of the profession are enmeshed in outmoded or at least questionable concepts of human behavior, particularly, the *assumption* that mental abilities are relatively immutable even with extensive and prolonged experience or intervention, and, furthermore, that they have almost overpowering genetic determinants.

It is my contention, nevertheless, that intelligence, if seen in a proper light, is worthy of continued serious study, and that this study can have increasing relevance to problems of schooling and educational policy. The performances required on many types of mental ability tests—tests of language competence, of ability to manipulate abstract concepts and rela-

tionships, of ability to apply knowledge to the solution of problems, and even of the ability to make simple and rapid comparisons of stimuli (as in a test of perceptual speed)—have great and obvious resemblances to performances required in school learning, and indeed in many other fields of human activity. If these performances are seen as based on learned, developed abilities of a rather generalized character, it would frequently be useful to assess the extent to which an individual had acquired these abilities. This could be for the purpose of determining the extent to which these abilities would need to be improved to prepare the individual for further experiences or learning activities, or of determining what kinds and amounts of intervention might be required to effect such improvements. These determinations, however, would have to be based on more exact information than we now have concerning the effects of different types of learning experiences, including observation, practice, instruction, and so forth, on the improvement of these abilities. As matters stand now, we know very little about the parameters governing the growth of such individual attributes as language competence, reasoning ability, and speed of cognitive operations.

Even if performances on mental ability tests have at least some genetic determinants, as they probably do, they undoubtedly have substantial environmental determinants. Environmental variance is for the most part the only kind of variance that educators have a right to work with, but educators need to know the limits to which, and the conditions under which, this variance can be reduced. If the technological apparatus of factor analysis, test theory, and general experimental design can be directed toward the production of this kind of knowledge, and if this knowledge can be expressed and validated in a form that can be used by practitioners—a large order, perhaps, but feasible in principle—then the devoted efforts that have produced this apparatus will eventually have paid off.

REFERENCES

- Alexander, W. P. Intelligence, concrete and abstract. *British Journal of Psychology, Monograph Supplement*, 1935, 6(19).
- American Psychological Association. Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 1954, 51. (Supplement)
- American Psychological Association. *Standards for educational and psychological tests and manuals*. Washington, D.C.: Author, 1966.
- Anastasi, A. *Differential psychology: Individual and group differences in behavior* (3rd ed.). New York: Macmillan, 1958.
- Anderson, J. E. The limitations of infant and preschool tests in the measurement of intelligence. *Journal of Psychology*, 1939, 8, 351-379.
- Angoff, W. H. (Ed.). *The College Board admissions testing program: A technical report on research and development activities relating to the Scholastic Aptitude Test and*

- Achievement Tests*. New York: College Entrance Examination Board, 1971. (a)
- Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971. (b)
- Ash, P. The implications of the Civil Rights Act of 1964 for psychological assessment in industry. *American Psychologist*, 1966, **21**, 797-803.
- Baker, K. H. Item validity by the analysis of variance: An outline of method. *Psychological Record*, 1939, **3**, 242-248.
- Baxter, B. (Ed.). American Psychological Association, Task Force on Employment Testing of Minority Groups: Job testing and the disadvantaged. *American Psychologist*, 1969, **24**, 637-650.
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. *Differential aptitude tests*. New York: Psychological Corporation, 1947-1969.
- Binet, A. Perceptions d'enfants. *Revue Philosophique*, 1890, **30**, 582-611.
- Binet, A., & Simon, T. Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 1905, **11**, 191-244.
- Binet, A., & Simon, T. Le développement de l'intelligence chez les enfants. *L'Année Psychologique*, 1908, **14**, 1-94.
- Binet, A., & Simon, T. L'intelligence des imbéciles. *L'Année Psychologique*, 1909, **15**, 1-147.
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (Part 5). Reading, Mass.: Addison-Wesley, 1968.
- Bloom, B. S. *Stability and change in human characteristics*. New York: Wiley, 1964.
- Bloom, B. S. *Human characteristics and school learning*. New York: McGraw-Hill, 1976.
- Bôcher, M. *Introduction to higher algebra*. New York: Macmillan, 1907.
- Bock, R. D., & Wood, R. Test theory. *Annual Review of Psychology*, 1971, **22**, 193-224.
- Boring, E. G. Intelligence as the tests test it. *New Republic*, 1923, **35**, 35-37.
- Bridgman, P. W. *The logic of modern physics*. New York: Macmillan, 1927.
- Brigham, C. C. Intelligence tests of immigrant groups. *Psychological Review*, 1930, **37**, 158-165.
- Brim, O. G., Jr., Neulinger, J., & Glass, D. C. *Experiences and attitudes of American adults concerning standardized intelligence tests* (Tech. Rep. 1 on the Social Consequences of Testing). New York: Russell Sage Foundation, 1965.
- Brody, E. B., & Brody, N. *Intelligence: Nature, determinants, and consequences*. New York: Academic Press, 1976.
- Brown, W. Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 1910, **3**, 296-322.
- Brown, W., & Thomson, G. *The essentials of mental measurement*. Cambridge, England: Cambridge University Press, 1921.
- Buros, O. K. (Ed.). *The nineteen forty mental measurements yearbook* (2 vols.). Highland Park, N.J.: Mental Measurements Yearbook, 1941.
- Buros, O. K. (Ed.). *The third mental measurements yearbook*. New Brunswick, N.J.: Rutgers University Press, 1949.
- Buros, O. K. (Ed.). *The fourth mental measurements yearbook*. Highland Park, N.J.: Gryphon Press, 1953.
- Buros, O. K. (Ed.). *The fifth mental measurements yearbook*. Highland Park, N.J.: Gryphon Press, 1959.
- Buros, O. K. (Ed.). *The sixth mental measurements yearbook*. Highland Park, N.J.: Gryphon Press, 1965.

- Buros, O. K. (Ed.). *The seventh mental measurements yearbook*. Highland Park, N.J.: Gryphon Press, 1972.
- Buros, O. K. (Ed.). *Tests in print II: An index to tests, test reviews, and the literature on specific tests*. Highland Park, N.J.: Gryphon Press, 1974.
- Burt, C. Factor analysis by sub-matrices. *Journal of Psychology*, 1938, **6**, 339-375.
- Burt, C. The education of the young adolescent: Psychological implications of the Norwood Report. *British Journal of Educational Psychology*, 1943, **13**, 126-140.
- Campbell, N. R. *An account of the principles of measurement and calculation*. London: Longmans, 1928.
- Cancro, R. (Ed.). *Intelligence: Genetic and environmental influences*. New York: Grune & Stratton, 1971.
- Carroll, J. B. A factor analysis of verbal abilities. *Psychometrika*, 1941, **6**, 279-307.
- Carroll, J. B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, **10**, 1-19.
- Carroll, J. B. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23-38.
- Carroll, J. B. The nature of the data, or how to choose a correlation coefficient. *Psychometrika*, 1961, **26**, 347-372.
- Carroll, J. B. Factors in verbal achievement. In P. L. Dressel (Ed.), *Proceedings of the Invitational Conference on Testing Problems, 1961*. Princeton, N.J.: Educational Testing Service, 1962.
- Carroll, J. B. A model of school learning. *Teachers College Record*, 1963, **64**, 723-733.
- Carroll, J. B. Stalking the wayward factors: Review of J. P. Guilford & Ralph Hoepfner, *The analysis of intelligence*. *Contemporary Psychology*, 1972, **17**, 321-324.
- Carroll, J. B. Fitting a model of school learning to aptitude and achievement data over grade levels. In D. R. Green (Ed.), *The aptitude-achievement distinction: Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement*. Monterey, Calif.: CTB/McGraw-Hill, 1974.
- Carroll, J. B. Psychometric tests as cognitive tasks: A new "Structure of Intellect." In L. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Erlbaum, 1976.
- Cattell, J. M. Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philosophische Studien*, 1885, **2**, 635-650.
- Cattell, J. M. Psychometrische Untersuchungen. *Philosophische Studien*, 1886, **3**, 452-492. (a)
- Cattell, J. M. The time taken up by cerebral operations. *Mind*, 1886, **11**, 222-242, 377-392, 524-538. (b)
- Cattell, J. M. Mental tests and measurements. *Mind*, 1890, **15**, 373-381.
- Cattell, J. M., & Farrand, L. Physical and mental measurements of the students of Columbia University. *Psychological Review*, 1896, **3**, 618-648.
- Cattell, R. B. The measurement of adult intelligence. *Psychological Bulletin*, 1943, **45**, 153-193.
- Cattell, R. B. *Abilities: Their structure, growth, and action*. Boston: Houghton-Mifflin, 1971.
- Chauncey, H., & Dobbin, J. E. *Testing: Its place in education today*. New York: Harper & Row, 1963.
- Christie, T., & Griffin, A. The examination achievements of highly selective schools. *Educational Research*, 1970, **12**, 202-208.
- College Entrance Examination Board. *Report of the Commission on Tests: I. Righting the balance*. New York: Author, 1970.
- Conant, J. B. *The American high school today: A first report to interested citizens*. New York: McGraw-Hill, 1959.

- Cooley, W. W., & Lohnes, P. R. *Multivariate procedures for the behavioral sciences*. New York: Wiley, 1962.
- Coombs, C. H. A factorial study of number ability. *Psychometrika*, 1941, **6**, 161-189.
- Crano, W. D., Kenny, D. A., & Campbell, D. T. Does intelligence cause achievement? A cross-lagged panel analysis. *Journal of Educational Psychology*, 1972, **63**, 258-275.
- Crawford, A. B. *Incentives to study*. New Haven: Yale University Press, 1929.
- Cremin, L. A. *The transformation of the school: Progressivism in American education, 1876-1957*. New York: Knopf, 1961.
- Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297-334.
- Cronbach, L. J. Five decades of public controversy over mental testing. *American Psychologist*, 1975, **30**, 1-14.
- Cronbach, L. J., & Furby, L. How should we measure "change"—Or should we? *Psychological Bulletin*, 1970, **74**, 68-80.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley, 1972.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. Theory of generalizability: A liberation of reliability theory. *British Journal of Mathematical and Statistical Psychology*, 1963, **16**, 137-163.
- Davidson, W. M., & Carroll, J. B. Speed and level components in time limit scores: A factor analysis. *Educational and Psychological Measurement*, 1945, **5**, 411-427.
- Doolittle, M. H. Method employed in the solution of normal equations and the adjustment of a triangulation. *U.S. Coast and Geodetic Survey Report*, 1878, 115-120.
- Downey, M. T. *Carl Campbell Brigham: Scientist and educator*. Princeton, N.J.: Educational Testing Service, 1961.
- DuBois, P. H. *A history of psychological testing*. Boston: Allyn & Bacon, 1970.
- Dunlap, J. W. The Psychometric Society—Roots and powers. *Psychometrika*, 1942, **7**, 1-8.
- Dvorak, B. J. The new USES General Aptitude Test Battery. *Journal of Applied Psychology*, 1947, **31**, 372-376.
- Ebbinghaus, H. Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. *Zeitschrift für angewandte Psychologie*, 1897, **13**, 401-459.
- Ekstrom, R. B. *Cognitive factors: Some recent literature* (Project Rep. 73-30). Princeton, N.J.: Educational Testing Service, 1973.
- Ellison, M. L., & Edgerton, H. A. The Thurstone Primary Mental Abilities Tests and college marks. *Educational and Psychological Measurement*, 1941, **1**, 399-406.
- Eysenck, H. J. (Ed.). *The measurement of intelligence*. Baltimore: Williams & Wilkins, 1973.
- Fechner, G. *Elemente der Psychophysik*. Leipzig: Breitkopf & Härtel, 1860.
- Ferguson, G. A. Item selection by the constant process. *Psychometrika*, 1942, **7**, 19-29.
- Ferguson, G. A. On transfer and the abilities of man. *Canadian Journal of Psychology*, 1956, **10**, 121-131.
- Fincher, C. Is the SAT worth its salt? An evaluation of the use of the Scholastic Aptitude Test in the university system of Georgia over a thirteen-year period. *Review of Educational Research*, 1974, **44**, 293-305.

- Findley, W. G., & Bryan, M. M. *Ability grouping: 1970. Status, impact, and alternatives*. Athens, Ga.: University of Georgia, Center for Educational Improvement, 1971.
- Finney, D. J. The application of probit analysis to the results of mental tests. *Psychometrika*, 1944, **9**, 31-39.
- Fisher, R. A. *Statistical methods for research workers*. Edinburgh & London: Oliver & Boyd, 1925.
- Flanagan, J. C. (Ed.). *The aviation psychology program in the Army Air Forces* (Rep. 1, Army Air Forces Aviation Psychology Program Research Reports). Washington, D.C.: U.S. Government Printing Office, 1948.
- Flanagan, J. C. *Flanagan aptitude classification tests*. Chicago: Science Research Associates, 1951-1960.
- Flanagan, J. C. Truman Lee Kelley [obituary notice]. *Psychometrika*, 1961, **26**, 343-345.
- Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Gorham, W. A., Orr, D. B., & Goldberg, I. *Design for a study of American youth. 1: The talents of American youth*. Boston: Houghton-Mifflin, 1962.
- Flanagan, J. C., & Cooley, W. W. *Project talent: One year follow-up studies*. Pittsburgh: University of Pittsburgh, Project Talent, 1966.
- Flanagan, J. C., & Jung, S. M. *Progress in education: A sample survey, 1960-1970*. Palo Alto, Calif.: American Institutes for Research, 1971.
- French, J. L. (Reviser). *Henmon-Nelson Tests of Mental Ability, 1973 revision*. Boston: Houghton-Mifflin, 1973.
- French, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monographs*, 1951, **5**.
- French, J. W. *Kit of selected tests for reference aptitude and achievement factors*. Princeton, N.J.: Educational Testing Service, 1954.
- French, J. W., Ekstrom, R. B., & Price, L. A. *Kit of reference tests for cognitive factors*. Princeton, N.J.: Educational Testing Service, 1963.
- Fulker, D. W. Review of Kamin's *The science and politics of I.Q.* *American Journal of Psychology*, 1975, **88**, 505-519.
- Galton, F. *Hereditary genius: An inquiry into its laws and consequences*. New York: Appleton, 1869.
- Galton, F. *Inquiries into human faculty and its development*. London: Macmillan, 1883.
- Galton, F. Some results of the Anthropometric Laboratory. *Journal of the Anthropological Institute*, 1885, **14**, 275-287.
- Galton, F. Family likeness in stature. *Proceedings of the Royal Society of London*, 1886, **40**, 42-63.
- Galton, F. Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, 1889, **45**, 135-145.
- Goddard, H. H. The Binet and Simon tests of intellectual capacity. *The Training School*, 1908, **5**, 3-9.
- Gordon, E. W., & Wilkerson, D. A. *Compensatory education for the disadvantaged*. New York: College Entrance Examination Board, 1966.
- Goslin, D. A. *The search for ability: Standardized testing in social perspective*. New York: Russell Sage Foundation, 1963.
- Green, B. F., Jr. The computer revolution in psychometrics. *Psychometrika*, 1966, **31**, 437-445.
- Green, D. R. (Ed.). *The aptitude-achievement distinction: Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement*. Monterey, Calif.: CTB/McGraw-Hill, 1974.

- Green, D. R., Ford, M. P., & Flamer, G. B. (Eds.). *Measurement and Piaget*. New York: McGraw-Hill, 1971.
- Guilford, J. P. (Ed.). *Printed classification tests* (Res. Rep. 5, Army Air Forces, Aviation Psychology Program). Washington, D.C.: Government Printing Office, 1947.
- Guilford, J. P. The structure of intellect. *Psychological Bulletin*, 1956, **53**, 267-293.
- Guilford, J. P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.
- Guilford, J. P., & Hoepfner, R. *The analysis of intelligence*. New York: McGraw-Hill, 1971.
- Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman Aptitude Survey*. Beverly Hills, Calif.: Sheridan Supply Co., 1947-1950.
- Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.
- Guttman, L. Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 1940, **5**, 75-99.
- Guttman, L. The quantification of a class of attributes: A theory and method for scale construction. In P. Horst (Ed.), *The prediction of personal adjustment*. New York: Social Science Research Council, 1941.
- Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika*, 1953, **18**, 277-296.
- Guttman, L. Reliability formulas for non-completed or speeded tests. *Psychometrika*, 1955, **20**, 113-124.
- Hakstian, A. R., & Cattell, R. B. *Comprehensive Abilities Battery*. Champaign, Ill.: Institute for Personality and Ability Testing, 1976.
- Harman, H. H., Ekstrom, R. B., & French, J. W. *Kit of factor reference cognitive tests*. Princeton, N.J.: Educational Testing Service, 1976.
- Harnischfeger, A., & Wiley, D. E. *Achievement test score decline: Do we need to worry?* Saint Louis, Mo.: CEMREL, 1976.
- Helmholtz, H. von. Zahlen und Messen, erkenntniss- heoretisch betrachtet. In H. von Helmholtz, *Wissenschaftliche Abhandlungen; III*. Leipzig: Barth, 1885.
- Henmon, V. A. C., & Nelson, M. J. *Henmon-Nelson test of mental ability*. Boston: Houghton-Mifflin, 1932-1935.
- Herrnstein, R. I. *I.Q. in the meritocracy*. Boston: Little, Brown, 1973.
- Hills, J. R. Use of measurement in selection and placement. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971.
- Hofstadter, R. *Social Darwinism in American thought*. Philadelphia: University of Pennsylvania Press, 1944.
- Hölder, L. O. Die Axiome der Quantität und die Lehre vom Mass. *Berichte, Sächsische Akademie der Wissenschaften, Math. nat. Klasse*, 1901, **53**, 1-64.
- Holmen, M. G., & Docter, R. *Educational and psychological testing: A study of the industry and its practices*. New York: Russell Sage Foundation, 1972.
- Horn, J. L. On subjectivity in factor analysis. *Educational and Psychological Measurement*, 1967, **27**, 811-820.
- Horn, J. L. Organization of abilities and the development of intelligence. *Psychological Review*, 1968, **75**, 242-259.
- Horn, J. L. Human abilities: A review of research and theory in the early 1970's. *Annual Review of Psychology*, 1976, **27**, 437-485.
- Horn, W. O. *International Primary Factors Test Battery*. Stevens Point, Wis.: International Tests, 1973.
- Horst, P. Relations among m sets of measures. *Psychometrika*, 1961, **26**, 129-149.
- Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 1933, **24**, 417-441, 498-520.

- Hotelling, H. The most predictable criterion. *Journal of Educational Psychology*, 1935, **26**, 139-142.
- Hoyt, C. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153-160.
- Humphreys, L. G. The misleading distinction between aptitude and achievement tests. In D. R. Green (Ed.), *The aptitude-achievement distinction: Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement*. Monterey, Calif.: CTB/McGraw-Hill, 1974.
- Hunt, E. Varieties of cognitive power. In L. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Erlbaum, 1976.
- Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *Advances in learning and motivation* (Vol. 7). New York: Academic Press, 1973.
- Hunt, J. M. *Intelligence and experience*. New York: Ronald Press, 1961.
- Husén, T. The concept of intelligence in applied psychology: Some introductory observations. *International Review of Applied Psychology*, 1975, **24**, 83-84.
- Jäger, A. O. *Dimensionen der Intelligenz*. Göttingen: Verlag für Psychologie, Hogrefe, 1967.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books, 1972.
- Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, **39**, 1-123.
- Jensen, A. R. *Genetics and education*. New York: Harper & Row, 1972.
- Jensen, A. R. *Educability and group differences*. London: Methuen, 1973. (a)
- Jensen, A. R. *Educational differences*. London: Methuen, 1973. (b)
- Jöreskog, K. G., & Lawley, D. N. New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1968, **21**, 85-96.
- Jones, L. V. The nature of measurement. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971.
- Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **2**, 187-200.
- Kaiser, H. F. Varimax solution for Primary Mental Abilities. *Psychometrika*, 1960, **25**, 153-158.
- Kaiser, H. F., & Caffrey, J. Alpha factor analysis. *Psychometrika*, 1965, **30**, 1-14.
- Kamin, L. J. *The science and politics of I.Q.* Potomac, Md.: Erlbaum, 1974.
- Keliher, A. V. *A critical study of homogeneous grouping* (No. 452, Contributions to Education). New York: Columbia University, Teachers College, 1931.
- Kelley, T. L. *Statistical methods*. New York: Macmillan, 1923.
- Kelley, T. L. *The interpretation of educational measurements*. Yonkers-on-Hudson, N.Y.: World Book, 1927.
- Kelley, T. L. *Crossroads in the mind of man: A study of differentiable mental abilities*. Stanford, Calif.: Stanford University Press, 1928.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement* (Vol. 1). New York: Academic Press, 1971.
- Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151-160.
- Labov, W. The logic of nonstandard speech. *Georgetown University Monographs of Languages and Linguistics*, 1969, **22**, 1-43.
- Lambert, C. M. Symposium on education of pupils for different types of secondary schools. VII: A survey of ability and interest at the stage of transfer. *British Journal of Educational Psychology*, 1949, **19**, 67-81.

- Lamke, T. A., & Nelson, M. J. *The Henmon-Nelson Tests of Mental Ability* (Rev. ed.). Boston: Houghton-Mifflin, 1957-1958.
- Levy, P. On the relation between test theory and psychology. In P. Kline (Ed.), *New approaches in psychological measurement*. London & New York: Wiley, 1973.
- Linden, K. W., & Linden, J. D. *Modern mental measurement: A historical perspective*. Boston: Houghton-Mifflin, 1968.
- Linn, R. L. A Monte Carlo approach to the number of factors problem. *Psychometrika*, 1968, **33**, 37-72.
- Loehlin, J. C., & Nichols, R. C. *Heredity, environment, and personality: A study of 850 sets of twins*. Austin: University of Texas Press, 1976.
- Loevinger, J. A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs*, 1947, **61**(4).
- Loevinger, J. The attenuation paradox in test theory. *Psychological Bulletin*, 1954, **51**, 493-504.
- Loevinger, J., Gleser, G. C., & DuBois, P. H. Maximizing the discriminating power of a multiple-score test. *Psychometrika*, 1953, **18**, 309-317.
- Lord, F. M. A theory of test scores. *Psychometric Monographs*, 1952, **7**.
- Lord, F. M. Some perspectives on "the attenuation paradox" in test theory. *Psychological Bulletin*, 1955, **52**, 505-510.
- Lord, F. M. A study of speed factors in tests and academic grades. *Psychometrika*, 1956, **21**, 31-50.
- Lord, F. M. A theoretical study of two-stage testing. *Psychometrika*, 1971, **36**, 227-242.
- Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.
- Loretan, J. O. The decline and fall of group intelligence testing. *Teachers College Record*, 1965, **67**, 10-17.
- Lumsden, J. Test theory. *Annual Review of Psychology*, 1976, **27**, 251-280.
- Marjoribanks, K. Environment as a threshold variable: An examination. *Journal of Educational Research*, 1974, **67**, 610-616.
- McClure, W. E. The status of psychological testing in large city public school systems. *Journal of Applied Psychology*, 1930, **14**, 486-496.
- McCullough, C. M. Relationship between intelligence and gains in reading ability. *Journal of Educational Psychology*, 1939, **30**, 688-692.
- McDonald, R. P. A general approach to nonlinear factor analysis. *Psychometrika*, 1962, **27**, 397-415.
- Mehler, J., & Bever, T. G. Editorial. *Cognition*, 1973, **2**, 7-11.
- Meili, R. Die faktorenanalytische Interpretation der Intelligenz. *Schweizerische Zeitschrift für Psychologie*, 1964, **13**, 135-155.
- Monroe, W. S. Educational measurement in 1920 and 1945. *Journal of Educational Research*, 1945, **38**, 334-340.
- Mort, P. R., & Featherstone, W. B. The general uses of psychological tests. *Review of Educational Research*, 1932, **5**, 300-307.
- Norwood, C. *Curriculum and examinations in secondary schools: Report of the Committee of the Secondary School Examinations Council appointed by the President of the Board of Education in 1941*. London: H. M. Stationery Office, 1943.
- Novick, M. R. The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 1966, **3**, 1-18.
- Otis, A. S. An absolute point scale for the group measure of intelligence. *Journal of Educational Psychology*, 1918, **9**, 238-261.
- Otis, A. S. The Otis Correlation Chart. *Journal of Educational Research*, 1923, **8**, 440-448.

- Pearson, K. *The grammar of science*. London: Walter Scott, 1892.
- Pearson, K. Contributions to the mathematical theory of evolution. I: On the dissection of asymmetrical frequency curves. *Philosophical Transactions* (Ser. A), 1894, **185**, 71-110.
- Pearson, K. Mathematical contributions to the theory of evolution. III: Regression, heredity and panmixia. *Philosophical Transactions* (Ser. A), 1896, **187**, 253-318.
- Peterson, J. *Early conceptions and tests of intelligence*. Yonkers-on-Hudson: World Book, 1925.
- Pintner, R. *Intelligence testing: Methods and results* (2nd ed.). New York: Henry Holt, 1931.
- Pruzek, R. M., & Coffman, W. E. *A factor analysis of the mathematical sections of the Scholastic Aptitude Test* (College Entrance Examination Board Research and Development Rep. 65-6, No. 10, and Research Bulletin 66-12). Princeton, N.J.: Educational Testing Service, 1966.
- Putnam, H. Reductionism and the nature of psychology. *Cognition*, 1973, **2**, 131-146.
- Rankin, P. T. Pupil classification and grouping. *Review of Educational Research*, 1931, **1**, 200-230.
- Rasch, G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielson & Lydiche, 1960.
- Ravitch, D. The I.Q. Myth—Criticisms, complexities, contradictions. *New York Times*, April 20, 1975, Sec. D, p. 29.
- Rayburn, W. G., & Hayes, E. J. Compensatory education: Effective or ineffective? *Journal of Counseling Psychology*, 1975, **22**, 523-528.
- Resnick, L. B. (Ed.). *The nature of intelligence*. Hillsdale, N.J.: Erlbaum, 1976.
- Richardson, S. C. Some evidence relating to the validity of selection for grammar schools. *British Journal of Educational Psychology*, 1956, **26**, 15-24.
- Royce, J.-R. (Ed.). *Multivariate analysis and psychological theory*. London & New York: Academic Press, 1973.
- Saunders, D. R. *An analytic method for rotation to orthogonal simple structure* (Research Bulletin 53-10). Princeton, N.J.: Educational Testing Service, 1953.
- Scarr-Salapatek, S. Review of Kamin's *The science and politics of I.Q.* *Contemporary Psychology*, 1976, **21**, 98-99.
- Schmid, J., & Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika*, 1957, **22**, 53-61.
- Sharp, S. E. Individual psychology: A study in psychological method. *American Journal of Psychology*, 1898-1899, **10**, 329-391.
- Simon, B. *The politics of educational reform, 1920-1940*. London: Lawrence & Wishart, 1974.
- Smith, B. O. *Logical aspects of educational measurement*. New York: Columbia University Press, 1938.
- Spearman, C. The proof and measurement of association between two things. *American Journal of Psychology*, 1904, **15**, 72-101. (a)
- Spearman, C. "General intelligence," objectively determined and measured. *American Journal of Psychology*, 1904, **15**, 201-293. (b)
- Spearman, C. Correlation calculated from faulty data. *British Journal of Psychology*, 1910, **3**, 71-295.
- Spearman, C. *The nature of "intelligence" and the principles of cognition*. London: Macmillan, 1923.
- Spearman, C. *The abilities of man*. London: Macmillan, 1927.
- Spearman, C., & Jones, L. W. *Human ability: A continuation of "The abilities of man."*

- London: Macmillan, 1950.
- Stanley, J. C. Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika*, 1961, **26**, 205-219.
- Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971.
- Stanley, J. C., & Wang, M. D. Weighting test items and test-item options: An overview of the analytical and empirical literature. *Educational and Psychological Measurement*, 1970, **30**, 21-35.
- Stern, W. *Psychologische Methoden der Intelligenz-Prüfung*. Leipzig: Barth, 1912.
- Sternberg, R. J. *The componential analysis of human abilities: Intelligence, information processing, and analogical reasoning*. Unpublished Ph.D. dissertation, Stanford University, 1975.
- Stevens, S. S. On the theory of scales of measurement. *Science*, 1946, **103**, 667-680.
- Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
- Sullivan, E. T. Psychographic representation of results of the Stanford Revision of the Binet-Simon tests. *Journal of Delinquency*, 1926, **10**, 284-285.
- Sullivan, E. T., Clark, W. W., & Tiegs, E. W. *California Test of Mental Maturity*. Los Angeles: California Test Bureau, 1936-1939.
- Suppes, P., & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.
- Svensson, N. *Ability grouping and scholastic achievement: Report on a five-year follow-up study in Stockholm*. Uppsala: Almqvist & Wiksell, 1962.
- Tatsuoka, M. M. Multivariate analysis in educational research. *Review of Research in Education*, 1973, **1**, 273-319.
- Taylor, C. W. A factorial study of fluency in writing. *Psychometrika*, 1947, **12**, 239-262.
- Taylor, W. L. "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 1953, **30**, 415-433.
- Terman, L. M. *The measurement of intelligence*. Boston: Houghton-Mifflin, 1916.
- Terman, L. M., & Merrill, M. A. *Measuring intelligence*. Boston: Houghton-Mifflin, 1937.
- Terman, L. M., & Merrill, M. A. *Stanford-Binet intelligence scale*. Boston: Houghton-Mifflin, 1960.
- Thorndike, E. L. *Introduction to the theory of mental and social measurements*. New York: Columbia University, Teachers College, 1904.
- Thorndike, E. L., et al. Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, 1921, **12**, 123-147, 195-216.
- Thorndike, E. L., Bregman, E. O., Cobb, M. V., Woodyard, E., & the staff of the Division of Psychology of the Institute of Educational Research of Teachers College, Columbia University. *The measurement of intelligence*. New York: Columbia University, Teachers College, Bureau of Publications, 1927.
- Thorndike, R. L. *Personnel selection: Test and measurement techniques*. New York: Wiley, 1949.
- Thorndike, R. L. Intellectual status and intellectual growth. *Journal of Educational Psychology*, 1966, **57**, 121-127.
- Thorndike, R. L., & Hagen, E. *10,000 careers*. New York: Wiley, 1959.
- Thurstone, L. L. *The nature of intelligence*. London & New York: Harcourt, Brace, 1924.
- Thurstone, L. L. *Fundamentals of statistics*. New York: Macmillan, 1925. (a)

- Thurstone, L. L. A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 1925, **16**, 433-451. (b)
- Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927, **34**, 273-286.
- Thurstone, L. L. The absolute zero in intelligence measurement. *Psychological Review*, 1928, **35**, 175-197.
- Thurstone, L. L. *The reliability and validity of tests*. Ann Arbor, Mich.: Edwards Brothers, 1931. (a)
- Thurstone, L. L. Multiple factor analysis. *Psychological Review*, 1931, **38**, 406-427. (b)
- Thurstone, L. L. *The vectors of the mind*. Chicago: University of Chicago Press, 1935.
- Thurstone, L. L. Primary mental abilities. *Psychometric Monographs*, 1938, **1**. (a)
- Thurstone, L. L. *Tests for primary mental abilities: Experimental edition*, 1938. Washington, D.C.: American Council on Education, 1938. (b)
- Thurstone, L. L. Current issues in factor analysis. *Psychological Bulletin*, 1940, **37**, 189-236.
- Thurstone, L. L. *A factorial study of perception*. Chicago: University of Chicago Press, 1944.
- Thurstone, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.
- Thurstone, L. L. *The differential growth of mental abilities* (Psychometric Laboratory Rep. No. 14). Chapel Hill: University of North Carolina, 1955.
- Thurstone, L. L., & Thurstone, T. G. The 1935 psychological examination. *Educational Record*, 1936, **17**, 296-317.
- Thurstone, L. L., & Thurstone, T. G. *SKA Primary Mental Abilities*. Chicago: Science Research Associates, 1946-1953.
- Thurstone, L. L., Thurstone, T. G., & Adkins, D. C. The 1938 psychological examination. *Educational Record*, 1939, **20**, 263-300.
- Toops, H. A. Eliminating the pitfalls in solving correlation—A printed correlation form. *Journal of Experimental Psychology*, 1921, **4**, 434-446.
- Toops, H. A. The status of university intelligence tests in 1923-24. *Journal of Educational Psychology*, 1926, **17**, 22-36, 110-124.
- Tucker, L. R. Maximum validity of a test with equivalent items. *Psychometrika*, 1946, **11**, 1-13.
- Tucker, L. R. Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.
- Tyler, R. W., & Wolf, R. M. (Eds.). *Crucial issues in testing*. Berkeley, Calif.: McCutchan, 1974.
- Vandenberg, S. G. The hereditary abilities study: Hereditary components in a psychological test battery. *American Journal of Human Genetics*, 1962, **14**, 220-237.
- Vandenberg, S. G. Genetic factors in human learning. *Educational Psychologist*, 1976, **12**, 59-63.
- Wade, N. I.Q. and heredity: Suspicion of fraud beclouds classic experiment. *Science*, 1976, **194**, 916-919.
- Walker, D. A. Answer pattern and score scatter in tests and examinations. *British Journal of Psychology*, 1931, **22**, 73-86.
- Walker, D. A. Answer pattern and score scatter in tests and examinations. *British Journal of Psychology*, 1936, **26**, 301-308.

- Walker, D. A. Answer pattern and score scatter in tests and examinations. *British Journal of Psychology*, 1940, **30**, 248-260.
- Walker, H. M. *Studies in the history of statistical method*. Baltimore: Williams & Wilkins, 1929.
- Warren, R., & Mendenhall, R. M. *The Mendenhall-Warren-Hollerith correlation method*. New York: Columbia University Statistics Bureau, 1929.
- Werdelin, I. *Geometrical ability and the space factor in boys and girls*. Lund, Sweden: Gleerups, 1961.
- Werts, C. E., & Linn, R. L. Path analysis: Psychological examples. *Psychological Bulletin*, 1970, **74**, 193-212.
- Wherry, R. J. A modification of the Doolittle method: A logarithmic solution. *Journal of Educational Psychology*, 1932, **23**, 455-459.
- Wherry, R. J., & Gaylord, R. H. The concept of test and item reliability in relation to factor pattern. *Psychometrika*, 1943, **8**, 247-264.
- Wherry, R. J., & Winer, B. J. A method for factoring large numbers of items. *Psychometrika*, 1953, **18**, 161-179.
- Whipple, G. M. *Manual of mental and physical tests*. Baltimore: Warwick & York, 1910.
- Wissler, C. The correlation of mental and physical traits. *Psychological Monographs*, 1901, **3**, 1-62.
- Wolfe, D. Factor analysis to 1940. *Psychometric Monographs*, 1940, **3**.
- Wood, B. L. *Measurement in higher education*. Yonkers: World Book, 1923.
- Wright, S. Correlation and causation. *Journal of Agricultural Research*, 1921, **20**, 557-585.
- Wundt, W. *Beiträge zur Theorie der Sinneswahrnehmung*. Leipzig & Heidelberg: Winter, 1862.
- Yerkes, P. M. (Ed.). Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences*, 1921, **15**.
- Young, M. *The rise of the meritocracy, 1870-2033: An essay on education and equality*. London: Thames & Hudson, 1958.
- Yule, G. U. On the theory of correlation. *Journal of the Royal Statistical Society*, 1897, **60**, 812-854.
- Yule, G. U. *An introduction to the theory of statistics*. London: Griffin, 1911.

Chapter 2

WORDS FOR SCHOOLS: THE APPLICATIONS IN EDUCATION OF THE VOCABULARY RESEARCHES OF EDWARD L. THORNDIKE

by

GERALDINE JONCICH CLIFFORD

University of California
at Berkeley

Each word which he [the child] learns controls for the moment his thought. Each sentence that he hears shapes for him a succession of ideas and holds them in a certain order. After the individual has been under the influence of language for a time, he will begin to do most of his thinking by using the distinctions and relations which society has given him through its established modes of communication. Language is a mode of mental procedure. It is not something which we use from time to time; it is the method of our whole mental existence. An individual can no more get away from language than he can empty his mind of ideas. (Judd, 1927, p. 208)

NO IRONY is intended by introducing this chapter, on the contributions of Edward Lee Thorndike to the language of the school and schoolbook, with the words of Charles Hubbard Judd—widely regarded as Thorndike's closest rival as America's preeminent educational psychologist of the first third or more of the 20th century. What is intended is to suggest that a serious interest in words united the representatives and followers of differing psychological and philosophical traditions in modern American education.

Appreciation is extended to Jeanne S. Chall, John B. Carroll, and Herbert Simons who offered valuable suggestions for the research of this essay; they are not, of course, responsible for any errors of omission or commission. Thanks are also due to Sonya Kaufman for general reference assistance and to W. F. Clifford who researched the biographical directories, especially the volumes of *Leaders in Education*, to verify the personal linkages between vocabulary researchers.

Indeed, "vocalarists" were wont to quote from John Dewey's *How We Think* (1910), especially when their word counts were shrugged off as "words, mere words." Pursuing what linguist Leonard Bloomfield (1933, p. 500) dismissed as a "metaphysical doctrine," laymen and professionals continued to ponder the relationship between one's stock of words and one's stock of ideas (Seago, 1925). Edwin Doran (1907) was drawn to vocabulary research by the contention in a 19th-century grammar book that "a word is a sign of an idea" (p. 431). "Of all the inventions of the human race nothing compares in importance, as regards mental development, with language," enthused a professor of the Fitchburg, Massachusetts, State Normal School (Kirkpatrick, 1907, p. 157). Words are the "raw materials not only of reading but also of thinking," wrote Luella Cole (1934, p. 47) in a psychology textbook for prospective teachers. At a conference of educators, Edgar Dale (1947) quoted the boy who said, "I don't know what I think about that yet. I haven't written anything on it," and the girl who said, "I talk so as to find out what I think—don't you?" (p. 94).

Edward L. Thorndike (1874–1949) has to be counted a notable exception to Bloomfield and Mowrer's observation that "psychologists generally treat language as a side issue" (Bloomfield, 1933, p. 512; Mowrer, 1954, p. 660). Thorndike's publications specific to language date from 1908 to 1949 and number around 75. The plan of this chapter is, first, to account for the rising interest in, and susceptibility to, scientific vocabulary research, as older traditions of vocabulary investigation interacted with new scientific and cultural movements; second, to place in its historical context the dominant position in vocabulary research earned by Thorndike's *The Teacher's Word Book* and indicate the maturing of a tradition that survives to the present; third, to describe the major "findings" of word-frequency research, their implications for educational theory and practice, and some of their certain and probable consequences in school operations; and, fourth, to complete the documentation of the direct and indirect roles played by *The Teacher's Word Book* in making the materials written for school use more readable, with special attention paid to the different internal characteristics of the several curriculum fields.

The adequacy, elegance, rigor, or definitiveness of vocabulary and readability research is not the issue here. Nor is the evidence at hand that would establish whether, on balance, there were more "good" outcomes than "bad" outcomes as a result of accepting some of these researchers' findings and the heightened pedagogical awareness of vocabulary issues to which they contributed. Nor is it demanded of impact data that they show that the learning problems of students of 1920 have been forestalled or erased by word research (Otto, 1971, p. 439). Rather, the purpose here is to describe and explain what happened in the word-world of elementary and secondary schooling in the period from 1920 to 1970

and to cast some light on the factors promoting, retarding, or moderating educational innovation.

Attention will be paid to a line of criticism frequently directed at educational research in general, considering the question of whether or not—on theoretical, methodological, or developmental grounds—vocabulary and readability research represents anything of an exception. Clifford (1973, pp. 34–35) has argued that much educational research is fragmented in character and ignorant of its own history; a cumulative research tradition, however, stands more chance of competing successfully for the schools' attention and seeing its findings influence educational practice. Writing of the mass of reading research, Fries (1963, p. 3) faults it for lack of "cumulative continuity," for not having those "strands of fundamental assumptions and accepted criteria of sound procedures" essential to good science. Klare (1963), himself a contributor to readability research and thus an heir of vocabulary studies, comments on its essentially empirical character: "If factors or methods, however arrived at, have worked well, they have been retained, and if not, they have been discarded" (pp. 182–183). Where one interpreter of readability research sees the healthy processes of incorporation, expansion, refinement, and correction at work—among men and women who were one another's teachers, students, and critics—Klare appears to emphasize the impoverishment of a research tradition in search of an adequate theoretical framework.

1. THE ORIGINS OF CONTEMPORARY INTEREST IN OBJECTIVE VOCABULARY DATA

1.1 *The Issues of Adult Vocabulary Size*

Amateurs ordinarily preceded professionals in areas of scientific investigation. In vocabulary studies, learned men—in the gentleman-scholar tradition—counted the running and different words variously found in Shakespeare, Milton, and the Bible. Fascinated with their own language behavior, they sought comparisons with individuals of different gender or occupational and educational level. The size of average adult vocabulary interested them, as did questions of the rate of annual vocabulary growth and of growth to what age. Not surprisingly—considering the expansion of knowledge in the later 19th century and the consequent increase of specialization in scholarship, science, and the professions—estimations also compared the general (common) with the technical (uncommon) lexicon.

In an oft-cited study reported in *Science* in 1891, E. A. Kirkpatrick concluded that "it is probable that to read ordinary general reading in English understandingly one needs to be familiar with six to seven thousand words" (p. 108). Using his own and earlier counts, Kirkpatrick

hypothesized that adults of ordinary intelligence and common (elementary) school backgrounds averaged 10,000-word vocabularies, while college graduates knew upwards of 20,000 words. Vocabulary extent seldom increases past age 30, Kirkpatrick concluded, and he later (1907) reasoned that the number of words added in adulthood is probably matched by the number forgotten (p. 160). Given subsequent rises in levels of completed schooling, Kirkpatrick's estimates compare favorably with present-day conclusions: The average vocabulary increases about 1,000 words per year between ages 3 and 20, and those who thereafter increase their word knowledge much beyond learning proper names "can be called 'academic' persons" (Joos, 1964, p. 206).

Between Kirkpatrick and Joos, however, there have been wildly discrepant findings; for college undergraduates Doran (1907) reported 19,000 words in average vocabularies and Hartmann (1946) reported 200,000! Variations in sampling techniques are a prime factor (Babbitt, 1907; Seashore, 1933; Gansl, 1939; Seegers & Seashore, 1949; Lorge & Chall, 1963). Kirkpatrick's (1907) method was to sample systematically an abridged dictionary, test himself on the words selected, and project the proportion known to the proportion the list was of the total number of words in the dictionary. He followed this method in a study of pupils from Grade 2 through the normal school and college levels. For a few years the debate centered upon dictionary size. Terman and Childs (1912) opted for an abridged version as a better test of the language's most basic words and as freer of cultural and educational biases. If the dictionary were not comprehensive, challenged Gillette (1929), vocabulary estimates of those with greater semantic and technical knowledge would be markedly underrated. With the appearance of Thorndike's *The Teacher's Word Book*, vocabulary counters had another selective source for their estimates: words ranked by frequency of occurrence in general reading materials. Its successor, *The Teacher's Word Book of 30,000 Words* (Thorndike & Lorge, 1944/1972), and random samples from an unabridged dictionary remain the chief sources of contemporary estimates of vocabulary size (Russell & Fea, 1963, p. 889).

During this period most research activities were assumed by professionals: psychologists and educators in the main, but also adult educators, public librarians, government bureaucrats, mass communications specialists, and military trainers. In 1907 Doran published a bibliography of 118 vocabulary studies and articles of varying substance and degrees of objectivity. Dale and Reichert (1957) cite 96 additional entries significantly concerned with vocabulary size and hundreds of others under related headings; 2,601 titles in vocabulary research were identified through 1955. Librarians campaigned for readable books for adults during the 1920s (Gray, 1937); the high, prolonged unemployment of the Great Depression further expanded library patronage and the concern for a

better matching of book and reader. Bureau of the Census data are suggestive: In 1920 only 28% of the nation's total population aged 14-17 were enrolled in the upper 4 years of secondary schools; in 1930 the figure was 47%. In 1940 the median highest school-grade level reached by American adults age 20 or over was reported as 8.4, and Gray and Leary (1935) used vocabulary-related calculations to estimate that one-half of published reading matter available in 1930 was too difficult to be read by half the adult population. As late as 1960, of the population aged 25 and older, 24 million Americans had completed less than 8 years of schooling. In the 1950s, armed forces reading programs were reported to be heavy users of research-derived readability formulas to select and grade books (Chall, 1956, p. 211).

The reading demands of naturalization—getting one's "citizen's papers"—slowly receded following the change in immigration policy, but the concern with readability of governmental publications—from tax forms and agricultural bulletins to driver's manuals—grew with the presence of ever more insistent and omnipresent governmental strictures. Chall (1956) and Klare (1963) indicate the uses made by business, industry, and the press in improving the readability of job applications, training materials, and the daily reading matter of the private sector. They, like the churches trying anew to reach more people, were hearing through "science" Paul's message (1 Corinthians 14:9): "Except ye utter by the tongue words easy to be understood, how shall it be known what is spoken?"

1.2 *The Vocabulary Development of the American Child*

In the 1830s, from the homes of educated and "enlightened" Europeans and Americans, came the first of a body of extant observations and recordings of children's language (McCarthy, 1933). Fueled by a rising compassion for child welfare and the sometimes highly sentimentalized fascination with "the estate of childhood" (Boas, 1966; Wishy, 1968), proud parents collected anecdotal language biographies in growing numbers; Clara and William Stern (he soon to become better remembered for originating the concept of the "intelligence quotient") summarized these in *Die Kindersprache* (Stern & Stern, 1907).

The child-study movement enlarged the enterprise after 1890 and concentrated the work in the United States under the enthusiastic guidance of G. Stanley Hall (Strickland & Burgess, 1965; Ross, 1972). Darwinism had spurred all sorts of developmental speculation and investigation, including the examination of language "evolution" in the child. With the rise of kindergartens, teachers joined parents in noting children's habits of oral language, knowledge of word meanings, and the progressive growth of reading and writing vocabularies. By the time Piaget's (1926) studies of children's thought processes, as they can be known

TABLE I

Estimates of Range of Vocabulary of the Average Pupil of Certain Grades or Mental Ages (Thorndike, 1924)

Grade	Kirkpatrick	Doran		Bonser						Brandenburg	Gerlach	Neher	Holly (Estimated)	Terman
		Dict. 1	Dict. 2	Dict. A	Dict. B									
				School A	School A	School B	School C	School D	School E					
2	4,480									4,000			3,144	M.A. 8 3,600
3	6,620									5,429			5,220	
4	7,020			10,886						6,887			6,030	M.A. 10 5,400
5	7,860			13,216	21,648	17,600	17,952	19,404	24,904	8,207			7,470	
6	8,700	4,192	11,634	17,038	30,184	26,520				9,613			8,478	M.A. 12 7,200
7	10,660	5,849	13,675	17,052	27,720					11,445			9,378	
8	12,000	7,937	17,138	18,704	31,120					12,819				
9	13,400		18,865											
1HS	15,640									13,504	66,650	9,000	10,296	M.A. 14 9,000
2HS	16,020									15,340		10,080	10,926	
3HS	17,600									13,974		11,520	11,565	Average Adult 11,700
4HS	18,720									14,975	76,250	11,700	11,752	Superior Adult 13,500
Normal School	19,000		15,750 19,400 21,500											
College Fresh.	20,120													
Soph.											73,650			
Jun.						35,376					85,300			
Sen.						37,092					86,925 94,075			

through language, were available in English, three types of American studies were available: (a) studies on the development of language through infancy and early childhood; (b) vocabulary lists of words known by children at various ages; and (c) studies of sentence formation, usage of the various parts of speech, and other elements of language complexity. The vocabulary studies were predominant until the appearance of psycholinguistics promoted studies in language acquisition, especially syntactic development (McCarthy, 1933, p. 329; Bellugi, 1964).

The first large, and most frequently cited, study of the recognition vocabularies of children was that of Madorah Smith (1926). A sample of words drawn from Thorndike's *The Teacher's Word Book* (1921/1927) was refined by comparisons with words used in 77 earlier children's vocabulary studies, including some of those summarized in Table 1. Objects, pictures of objects, and questions were used with 273 young children to elicit their knowledge of these words; projections led Smith to estimate that the average 4-year-old child knew about 1,500 words, his vocabulary growing to near 2,600 words at age 6.

Using a longitudinal study which followed the same children from kindergarten through the elementary school, Loban (1963) employed a vocabulary test that yielded a mean kindergarten vocabulary of 3,000 words known; individual children's vocabularies ranged from 180 words, of a principally Chinese speaking child, to about 5,000 words (p. 37). Given the increased opportunities for young children to learn words through television and travel, this modest growth from Madorah Smith's findings appeared plausible. Not so accepted were Mary K. Smith's results (1941). She reported growth in average vocabulary from 23,700 words (in Grade 1) to 80,300 (Grade 12). The grade-to-grade growth required by this larger estimate would mean that each day children gain knowledge of 10 new words; the conventional view was that about 1,000 new words are learned each school year by the average pupil (Dale, 1956, p. 122). Contemporary opinion is closer to the earlier, conservative estimates. Chall (1967), for example, writes, "The average *first grader* can probably use accurately and/or understand about four thousand different words" (p. 203).

As investigations of adult vocabularies clarified sampling issues, studies of children's language sharpened the question: What does it mean to "know" a word? Reviewing earlier studies of the vocabularies of children entering school, Dolch (1936) reminded researchers that they must be more careful in distinguishing the words children know from the words children use. As Hall (1883) had discovered, children commonly use words that have little or no meaning for them or to which they attach wrong meanings. The complicity of the schools in "verbalism" and all such "learning" without meaning was a major issue in the reform movement that came to be called progressive education (Cremin, 1961). Studies

(Chambers, 1904; Scott & Myers, 1923) and reviews (Colvin, 1951) detailed the problem. One common finding was pupils' mistaking an unknown word for one known: *unique* defined as "a man who lives in a harem," *belie* defined as "stomach," *abominable* as "stomach trouble" (Dale, 1956, p. 120).

Another crucial measurement issue was noted by Dolch (1936): specifying how much meaning shall be accepted as word knowledge. An investigator might ask respondents to pick out a picture or perform an action to show word knowledge, to supply a definition to a given word, to match given words and given definitions, to choose the correct from among alternative wrong answers ("distractors") of varying degrees of difficulty, or to choose the best of the several meanings of a given word to fit a given contextual requirement. A child who defines *policeman* as "a man in a blue uniform" or *oak* as "some kind of a tree" shows some unspecified familiarity with these words but less than concept mastery (p. 178). One of the functions of *The Teacher's Word Book* was to provide vocabulary test-makers with a tool other than sheer, unaided judgment for selecting word series that better judge the precision of a child's comprehension of the words being tested.

1.3 Vocabulary and the "New Type" Testing Movement

In December 1910, E. L. Thorndike appeared at a meeting of the American Federation of Teachers of Mathematics and Natural Sciences to talk of objective (quick-scoring), standardized tests. The subject aroused interest among geography teachers, as it had among other educational groups, and samples of the "new type" tests were reprinted in the *Journal of Geography* (Vol. 10) early in 1912.

Thorndike has been universally regarded as a figure of signal importance in the early development of school achievement and intelligence tests and scales; he was also a major contributor to the theory and methodologies that supported the broader measurement movement in education (Ayres, 1918; McCall, 1923; Garrison & Garrison, 1929; McCall, 1939; Cook, 1941; Wrightstone, Justman, & Robbins, 1956; Chauncey & Dobbin, 1963; Joncich, 1968). Thorndike, his early student B. R. Buckingham (1913), and Henry Rinsland (1935) were all test-makers at various times; each was also highly prominent among the word counters. They personified a coming together of vocabulary research and the testing movement.

In part, the relationship was an obvious one: Vocabulary researchers spent a great deal of time devising, or revising, or administering, or interpreting the results of the vocabulary tests being used to determine the word knowledge of various populations. Rough counts made of the titles entered under "Vocabulary Studies" in *The Education Index*—from its inception in 1929 through Volume 23 (1972–1973)—show that during the

period of most such activity, from 1929 to 1941, there was no year in which fewer than 10 such studies were published. While tests were not used in all this vocabulary research, they were surely employed in many.

Vocabulary tests, whether used by researchers or as part of a regular district-wide testing program, involved ever-larger numbers of teachers, students, and administrators after 1910. An average of four published vocabulary tests per year was reported in *The Education Index*; during the 9 years between 1935 and 1944, 76 new or revised versions appeared. Increasingly, the words for these tests came from one or a combination of the many word counts that appeared following the publication in 1921 of Thorndike's *The Teacher's Word Book* (Dale & Reichert, 1957). In most instances the makers of the new counts did not seek to surpass Thorndike's effort at a comprehensive frequency count of the general language of reading material. Rather, they sought to construct limited lists of specialized words that better reflected the technical vocabulary requirements of the several school subjects, or of reading materials more specific to children's interests, or as concerned speech or writing vocabularies—as differentiated from reading vocabularies. As early as 1930 there was hardly an area in the curriculum that lacked recourse to some pertinent list of "most important words." In some cases the lists were graded for school use, that is, arranged in order of relative difficulty by some procedure of validation.

Before educators began using standardized tests to assay their pupils' general and specialized word knowledge, psychologists had acquired experience at testing intelligence with measures of vocabulary (Seago, 1925). For one thing, there was that belief in virtual identity of language habits and mental processes already discussed. For another, as eight decades of experience have demonstrated, it is exceedingly difficult to construct nonverbal mental tests; hence, the typical intelligence scale forces careful consideration of word selection and arrangement. Additionally, test-makers were finding that a score made on the vocabulary subtests of their intelligence tests correlated most closely with the total score for the test (Terman, 1916, 1918; Smith, 1926; Kelley, 1928).

The intense discussions among psychologists about the nature of intelligence for the purpose of measurement provided vocabulary research with a new stimulus. Intelligence testing, indeed, the entire movement to develop and apply standardized tests, also provided—at least in its early years—vocabulary researchers with an attractive scientific and professional model.

1.4 Intelligence Tests and Pressure for Democratic Education

Intelligence tests were quickly perceived as measuring academic aptitude. Kirkpatrick (1907, p. 161) mentioned using their vocabulary sections in college entrance tests—this at a time, however, when America's

colleges were recruiting, not rejecting, even the less apt student. Scholastic aptitude tests furnished educators with more precise and sometimes more sobering information about their students than ever before available; standardized achievement tests did the same. Tests uncovered wide ranges of aptitude and achievement in a single classroom, within any given grade level, and between schools. In a large national sampling using tests of history vocabulary, Luella Cole found those great variations in performance that would be so often reported to teachers: "In Grade VIII the range is from almost complete failure to almost complete success" (Kelley & Krey, 1934, p. 165).

Mass education, through the grades. In 1922, standardized tests given in Iowa schools showed upper-term fifth-grader performances ranging over a 4-year span (Franzen & Knight, 1922, p. 84). By 1966, a textbook for social studies teachers reported that, by Grade 6, children's achievement in reading comprehension, vocabulary, arithmetic computation, and other school subjects varied by approximately 8 years (Hanna, Sabaroff, Davies, & Farrar, 1966, pp. 258-259). After accepting the possibility of some unreliability in the earlier study, but rejecting such alternative explanations as the exceptionality of Iowa children or marked changes up or down in the intellectual resources of the American population, two primary interpretations of these data do remain; they are not mutually exclusive in theory. One explanation is primarily pedagogical: American teaching has become progressively less effective than it formerly was with pupils of low motivation, from academically deprived backgrounds, and with limited capacities; and their opposite numbers learn and achieve largely in spite of the schools. The evidence to firmly establish or reject the truth, or even the partial truth, of this proposition is probably not even amenable of collection, much less of agreement of interpretation. The alternative thesis is chiefly political-structural: For internal and external reasons, the accepted pattern became that of classifying children primarily by chronological age-grading, extending upwards and enforcing compulsory attendance laws, and encouraging voluntary attendance through policies of "social (flexible) promotion," curricular variation, and official recognition of the student subculture.

That these changes did take place in American public schooling, to greater or lesser degree, varying in implementation from place to place and from one level of schooling to another, is freely reported and ordinarily documented—although large gaps in knowledge and evidence do remain (Coleman, 1961; Cremin, 1961; Clark, 1962; Krug, 1964, 1972; Rudy, 1965; Tyack, 1967; Wirth, 1972; Clifford, 1975). It seems fair to conclude, however, that the *effects* of such historical shifts, especially upon scholastic learning of the types measured by standardized tests, is not yet established even in outline form. What is unmistakable in the documentary record, however, is how much educators themselves perceived

changes as being necessitated by the commitment to mass, "democratic" education, even into the secondary schools. The question and answer were put thus by a civics teacher at Washington Irving High School in New York, the continuity of past and present educational thought being evident in the last sentence:

What shall we do with our low I.Q.'s? Make them welcome in our high schools and adapt our subject matter and technique to their needs. Try to lead them from the particular to the general, from the known to the unknown, at a gait adapted to their mental age and capacity. Choose for them a goal to which they can travel with a reasonable certainty of reaching it, though remote enough to make the journey alluring and the end worth while. . . . Try to substitute the question mark for the period in the minds of dull girls [and boys]. Above all, give them a moral code as a guidebook and train them in the use of it, that they may safely pass through the "storm and stress of youth" into the fruitful fields beyond. (Skinner, 1924, p. 228)

The early years of the educational research movement coincided with the maturing of the system-building begun in the 19th-century cities (Kaestle, 1973; Tyack, 1974; Troen, 1975). Insofar as the life of the teacher was concerned, its most bitter fruit grew from those reform efforts that abolished ungraded classrooms, once so common in both rural and urban areas, that spawned the insistent elaboration of courses of study and syllabi and promotion standards, and that codified rules of teacher and pupil behavior for whose enforcement an ever-increasing corps of supervisors appeared. Even before 1900, concern was voiced about pupils unable to conform to the standards of an increasingly rigid educational structure. Such students, now classified into grades, were forced to repeat as punishment for failure. The midyear promotion plan was one attempt to mitigate the effects of such nonpromotion (Bennett, 1917, p. 128). A few "special classes" were formed to recognize "the utter incapacity of some children to carry on the regular school work" (Mitchell, 1924, p. 6). Still, thousands of chronic repeaters piled up in the intermediate grades (Grades 4 to 6)—oversized, overage, and usually miserable—their own and their teacher's peace of mind depending upon chronic truancy and the promise of the inevitable "coming-of-age-14." Of children who are marked for repeated failure, Leonard Ayres (1909) wrote, "We know them in the schools as the children who are always a little behind physically, a little behind intellectually, and a little behind in the power to do. Such a child is the one who is always 'It' in the competitive games of childhood" (p. 220).

Data on the proportions of overage children in the elementary grades and disproportions of enrollments in the secondary schools allowed comparisons of the children of the foreign-born with children of native-born parentage (Miller, 1906); such data did not always favor the latter group but were nevertheless troubling to immigrant leaders. Public confidence

that tax-supported schools were either democratic or efficient was shaken by such studies as Thorndike's report, "The Elimination of Pupils from Schools" (1907), and Leonard Ayres's *Laggards in Our Schools* (1909). Of the children in school at age 8, Thorndike reported, only 30% remain to age 16 and 9% to age 18. Compared to the 90% who eventually satisfy the requirements for promotion from Grade 5, only 40% completed Grade 8. Writing of the nonpromoted, Thorndike concluded that "many pupils are held back unduly . . . [and] the work which they are given to do-but fail to do is unsuited to them" (p. 24).

Given the stimuli of fuller public information on the conditions within public education, the conversion of some teachers and administrators to the more humane and socially sensitive "new education" of Francis Parker and John Dewey, and (probably most important) the very great problems in classroom management that the "lock-step" mentality had encouraged, some modifications were introduced in promotion policies and attempts made to better fit instructional practices and school materials to the new facts about students. Nonetheless, in 1924 David Mitchell reminded the membership of the New York Society for the Experimental Study of Education that formal failure still confronted many New York City schoolchildren:

This year (1924) approximately two million children entered our country's schools for the first time. At the end of June 500,000 will remain in the first grade. Out of every four attempts to pass through the standard curriculum there will be one recorded failure. Even with promotions made as rapidly as possible, only 75 per cent of the children will be allowed to proceed with more advanced work. . . . Only one of every four who entered last fall will graduate from the eighth grade. Out of two million children, one and a half million will fall by the wayside. . . . We know that there are some children capable of more than twice the rate of progress or more than twice the amount of work which others can do. As all of them come under the supervision of our school authorities, the task is to find out what potentialities they have and then by experiment to make use of the ability and give to every child opportunities according to his mental equipment. (pp. 5, 16)

The roadblock of reading. During these years of educational ferment, Walter Percival undertook "A Study of the Causes and Subjects of School Failure" (1926) for a Teachers College doctoral dissertation. The survey identified reading disabilities as a major cause of pupil failure—from Grade 1, where the highest proportion of nonpromotions occurred in both rural and urban schools, through the succeeding grades; 68% of those not promoted in Grade 3 were kept back for this reason, as were 40% of fifth-grade and one-third of seventh-grade repeaters. Given low reading skill as a factor in difficulties encountered in the content fields and the probable, but difficult to measure, effect of poor reading skills on self-esteem, teacher attitude, and motivation to acquire knowledge through out-of-school activities, the influence of reading disability on

nonpromotion may have been even greater than Percival could report.

A growing corpus of research documented the observation that the "school is notoriously verbal" (MacGinitie, 1969, p. 695). Numerous tests of silent reading were prepared between 1910 and 1920 (Gray, 1938, pp. 102-103) and used to gain a more precise understanding of the relation between objectively measured reading achievement and such variables as intelligence, school marks, achievement in the several curriculum fields, avocational reading interests and habits, gender, social class, and ethnicity. Use of pupils' reading test scores came to supersede the mode of investigation used by Sholty (1912). Sholty's sample of second graders had been selected by teachers who had been asked simply to judge which were their best, poorest, and median-ability readers; moreover, her correlations between a pupil's word knowledge and his "general school standing" also relied on teacher judgment.

Studies correlating reading test scores and school marks were common in the early years of reading research (Gray, 1925, pp. 19-23); a positive correlation was regularly found. The research focus shifted somewhat, after 1925, as researchers used the new technical vocabulary lists to prepare test material; their interest was in the correlation of general reading ability (as determined by reading tests) and the several curriculum subjects. Such lists offered advantages, at least of efficiency, over the method of selecting arithmetic, history, or science terms from textbooks in use—textbooks that showed great vocabulary diversity. Given their lack of agreement on language usage and the dubious psychological and pedagogical principles shown in textbook construction, recourse to a competently developed, special vocabulary list was valued.

The intelligence testers had emphasized the contribution of word knowledge to the measurement of intelligence. There was, however, relatively little that schools could do with such knowledge except, perhaps, to use vocabulary tests as quick substitutes for intelligence tests in grouping students for instruction—a rather unpopular solution to the challenges of mass education. *More useful to the schools was knowing the role which "word power" played in reading, since reading was the widely acknowledged cause of so much personal pupil failure and shocking inefficiency in the system.* The findings of Hilliard (1924) were perhaps the most important in support of continued vocabulary research with more attention paid to the teaching of vocabulary. Of the several factors statistically determined as implicated in reading comprehension, he found that intelligence ranked first in importance and extent of vocabulary ranked second. Hilliard (1926) offered advice with his findings: "Training in vocabulary, with proper motivation, is a type of work well within the grasp of every teacher, and from this evidence seems to be one of the strategic points in building up comprehension" (p. 40).

In the second and third decades of this century, two fundamental

changes were effected in orthodox opinion on reading instruction. One was the shift of emphasis from oral to silent reading, even for children in the initial stages of learning to read. The second was the emphasis upon comprehension, which eclipsed other objectives of reading. If the conversion to the religion of silent and meaningful reading can be credited to any one man, he is William S. Gray.

Gray matriculated at Teachers College, Columbia University, immediately after earning a Bachelor of Science degree at the University of Chicago in 1913; his Columbia master's degree was earned in 1914. After completing this association with Thorndike, Gray returned to Chicago for the PhD (1916) and a long career as a renowned professor and dean—the author of influential books on the teaching of reading, highly regarded basal reading books for children, and authoritative periodic surveys and commentaries on reading research. In a brief remark in an extended description of preliminary work on what became Thorndike's reading and vocabulary scales and Gray's oral reading test, Thorndike (1914) reported having "cooperated with Mr. W. S. Gray in the construction of a first rough scale for measuring (1) simple, oral reading of matter-of-fact passages, and have devoted much time to scales for (2) silent reading so far as concerns (a) the understanding of words singly and (b) the understanding of sentences and paragraphs" (p. 208). Thorndike's prose style in the above precludes an attempt at a dramatic rendering of what appears to have been a historically significant personal association in the development of modern reading practices and educational research.¹

The goal of reading comprehension was symptomatic, and a special case, of a much more general reform movement in pedagogy—one which has acquired the title of the "progressive education movement." As well as being pedagogically progressive, both silent reading and reading for meaning were innovations that appealed to the utilitarian and realistic values in American culture; thus, they had cultural legitimacy from the beginning.

The *McGuffey Readers* are credited by historians with introducing several innovations which continued into the 20th century and were supported by vocabulary research: One such innovation was writing graded readers to keep pace with evolving graded schools; another was the principle of reducing the number of new words in each lesson and providing for their repetition (Chall, 1958b). But McGuffey ordinarily subordinated "sentence sense" to the presentation of and drill on the phonetic elements in words, his principal concern in primary reading instruction (Smith, 1965, p. 106). His choppy, isolated sentences failed

¹ The association of Gray and Thorndike merits further exploration given the probably exaggerated belief in the rivalry between the University of Chicago and Teachers College and the generalization that they represented polar opposites of educational and psychological thought.

one of the requirements of latter-day professional opinion, that word recognition and word analysis skills be taught in a meaningful narrative. Nor was McGuffey's use of literary selections, especially in intermediate-grade readers, a powerful example in the 1920s, when psychological principles became more evident in the selection of content. As Cronbach (1955) has explained, "Since 1900, for example, geometry texts have been drastically revised because of a new recognition that Euclid's form of exposition is not the most learnable [and] primary readers have been altered in manifold ways as it has become clear that just retelling 'worth-while' narratives is not the best way to build vocabulary and skill in purposeful reading" (p. 60).

To its adherents, "purposeful reading" means emphasis upon the ideas conveyed by words and by their relationship to one another, that is, "context." This conception of the association of reading with thinking was frequently defended by referring to Thorndike's celebrated studies of 1917 that defined the reading process as a reasoning process ("In Honor of Edward L. Thorndike," 1926, pp. 496-497; Russell, 1961; Otto, 1971; Stauffer, 1971; Tuinman, 1971). Reading for meaning must begin in the preprimer since emphasis in early reading upon phonetic elements, pronunciation and articulation drills, and word building from "meaningless" phonemes (akin to the psychologist's "nonsense syllables" as far as young readers are concerned) produces "mere word-callers"; while many children gained impressive proficiency in word-calling in the memoriter schools of the 19th century, such "verbalism" was scorned in the new century.

That the ability to pronounce words in print, even familiar words, should not be confused with reading comprehension was illustrated in the accumulating research. Fennell (1928) asked first graders and second graders to read and draw a picture to illustrate the sentence "In the circus there was a monkey with a yellow coat"; the children had previously responded correctly to other sentences using *with*. Some drew a monkey dragging a coat by its collar, sleeve, or tail; others drew a monkey and a coat. Only those children who had experienced (seen or seen pictured) a dressed monkey knew that one meaning of *with* is "dressed in"; even these children had to use the context to associate that meaning with their experience.

The conception of meaningful reading, then, requires that pedagogical attention be paid to children's background experiences. The most meaningful reading will be that which contains words and word relations already part of the student's recognition (hearing) vocabulary or which can be defined in such terms. The practice of having beginning readers share an experience, perhaps a walk to the fire house, then dictate a collective story of the experience, which the teacher writes down for the children and uses to further word recognition (i.e., building a *reading*

vocabulary), supposedly ensures that the words used in reading are well within both the children's recognition and oral (speech) vocabularies. Meaningful reading material, whether in experience charts or basal reading series, endeavors to tell stories or present information in reasonable relationship to a student's realm of experiences. The current appearance of basal reading and supplementary texts closer to the experiences of poor children and ethnic minorities conforms to such principles (Chall, 1967, pp. 44-49); today's concerns for relevance and realism, then, stem very much from this tradition.

In 1938, in *The Elementary English Review*, a basic journal in language teaching for teachers and supervisors, Donald Durrell and Helen Sullivan of the Educational Clinic of Boston University discussed the pressures that educators were transferring to publishers—that they produce more schoolbooks of simple vocabulary. They saw the need for easier reading material as the probable result of the change in school population caused by the lowering of school attrition over four decades: From the 90% who were lost between Grade 1 and the first year of high school in 1900 to a loss of less than 40% in 1938. Admitting that publishers of school materials were progressing in their simplifications, Durrell and Sullivan (1938) asked for still more precise information from textbook publishers about their practices of word selection: "There are many publishers who still dismiss the vocabulary burden of their textbooks by the simple statement, 'based on the Thorndike List', without specifying which levels of the Thorndike List are used" (p. 192). It is to the "Thorndike List" and its applications that we now turn.

2. STUDIES IN WORD USAGES: THE WORD COUNTERS

2.1 *Studies in Word Frequency Before 1921*

Word handbooks have been constructed on varying principles of word selection. One such criterion is word difficulty. The earlier English dictionaries were really glosses, *gloss* literally meaning "difficult," "obscure," "in need of explanation"; given this fact and the level of general education in early modern England, such dictionaries are more properly designated "scholars' handbooks." With greater occupational specialization, the technical importance of words emerged as another principle of word inclusion in a list. Even as early as the 15th century it was reflected in what eventually became the countless specialists' wordbooks—for apothecaries, dressmakers, electricians, or whatever.

The above exemplify "limited" lists, but the principle of comprehensiveness or all-inclusiveness was also used in gathering words together. The first successful effort to publish the complete English lexicon was that of Nathaniel Bailey in 1721. The successors of Bailey's "unabridged" dictionary have become so massive as not to qualify as handbooks, but

armloads. The very massiveness of modern dictionaries of the language and their inutility for such common purposes as a quick spelling reference for common words have maintained the need for abridged vocabularies. The basis of word selection in such abridgments is, when given, often very general. Although brief, the statement on word choice for the *Thorndike-Century Junior Dictionary* is specific:

Our selection of words is based on a count of the actual occurrences of words in over the million words of reading matter. These counts enable us to attach to each word a number showing its importance as measured by frequency and range of occurrence. The numbers 1 to 20 stand for successive thousands. A word having no number belongs in the third ten thousand. (Thorndike, 1935/1942, p. iii)²

By virtue of the research and care in the selection of entry items *and in the words used to define the entries*, this dictionary and its companions in the *Century* series (the *Senior* in 1941, the *Beginning* in 1945, and editions for England) were enormously successful (Barnhart, 1949). In the first 9 months after publication, the *Junior* averaged 1,000 copies sold per day; more than one million were sold in the first 9 years (Joncich, 1968, p. 576). A prime factor in its appeal was Thorndike's reputation among educators and librarians as the premier expert on the scientific selection of words. The publisher's advertisements for the *Junior Dictionary* headlined: "Publication of the THORNDIKE-CENTURY climaxes twenty-seven years of work, including compilation of the Teacher's Word Book." The success of a school dictionary based upon comprehensive counts of word frequency, was an inspiration for at least one other publisher since then, *The American Heritage School Dictionary* being based upon a very large computerized word-count project (Carroll, Davies, & Richman, 1971).

It was probably in 1907 that Thorndike began to attend seriously to word counts, to consider their possible utilities, perhaps even to begin some trials at counting. In this year, articles on vocabulary by Kirkpatrick, Babbitt, and Doran appeared in journals that Thorndike was likely to see. But much more specific to the compilation of vocabularies based on word frequency were the activities of two Teachers College colleagues: Gonzalez Lodge and Elijah Bagster-Collins.

A professor of Latin, Lodge published the results of work that he reportedly began in 1904 as *The Vocabulary of High School Latin* (Lodge, 1907; Lodge in Grinstead, 1924, p. 32). His is an arrangement of Latin words as found in the books that dominated introductory Latin courses,

² Thus, after the several definitions of *have* are given, in order of frequency of use, there appears the "importance number" of 1; *hatch* as in "to bring forth from an egg" has an importance of 5, but *hatch* as in "a divided door or cargo opening" is listed and defined without number—since that word of those meanings was not found frequently enough in reading matter to be included in *A Teacher's Word Book of Twenty Thousand Words* (1931/1932), upon which this dictionary was based.

words arranged first alphabetically and then in order of frequency of occurrence, beginning with *sum* and *omnis*. Lodge (1907) contended, of the most frequently used words on his list, that "a student who has at his command these 2000 words will have the vocabulary of fully nine tenths of all the ordinary Latin that he is likely to come into contact with" (p. v). He mentioned (p. iii) a similar effort at arranging Latin words on the frequency principle published by a Professor Fleckeisen nearly a century earlier in Germany and still in use there.

Elijah Bagster-Collins, professor of German and friend and neighbor of Thorndike's, has been credited with having been the first American known to have suggested, in 1908 and 1909, that frequency counts be used to standardize reading materials in teaching modern foreign languages (cited in Fries, 1950, p. 33n). Undoubtedly Bagster-Collins knew of the Lodge and German Latin lists. Probably he was aware of what was apparently known to leaders in foreign language instruction and what has become a classic in the history of vocabularies, F. W. Kaeding's *Häufigkeitwörterbuch der Deutschen Sprache* (Lorge, 1944; Fries, 1950). Published in 1898 and similar to certain much earlier word-frequency studies, it was written to assist shorthand writers. Based on an examination of nearly 11 million running words, in a wide range of materials, Kaeding's was a formidable achievement, a model for later word counters. Not as prominent, but suggesting that list-making had a general appeal, was the shorter list of 1,500 common Russian words, published in 1889 by N. A. Rubakin (Klare, 1963).

Early in 1908 Thorndike was reportedly at work writing a simplified primer of stories for slow readers; it did not become very popular, which Thorndike privately attributed to parental reluctance to admit the reality of low aptitude for learning (Joncich, 1968, p. 395). Nevertheless, by 1911 there was another flurry of activity testifying to Thorndike's growing involvement with the issue of the readability of school materials. His former student and colleague, Arthur Gates (1926), recalled that it was "about 1911, the year in which the course 'Psychology of the Elementary School Subjects' was first offered, that he [Thorndike] gave personal attention to the methods of evaluating materials. . . . Actual counts of arithmetic operations found in school texts, of words found in all sorts of reading materials, were soon undertaken" (p. 549). In a 1921 article about his word counts, Thorndike had indeed stated, "So for about ten years I have made such counts as I could" (p. 334). Also in 1911 came a modest frequency study of English vocabulary, R. C. Eldridge's word-frequency study of newspapers, *Six Thousand Common English Words*. The chief contribution of this list was the sharpened interest it generated in certain methodological issues that remained with researchers: the definition of what constitutes a word and what effects on reliability, validity, and durability the type of material sampled has (Lorge, 1944; Fries, 1950).

Aside from teachers of foreign languages, who already believed in the need to restrict the size of the vocabulary offered in early language instruction, educators appeared to think of word lists primarily as a spelling issue. The fact that the schools, despite long hours of instruction, were not producing good spellers got sensational press coverage after the exposés of Joseph Mayer Rice (1897). In 1913 Leonard Ayres published *The Spelling Vocabularies of Personal and Business Letters*. Ayres had wanted to use frequency as the criterion in establishing a reliable core of words for teaching in the lower grades, words whose spelling, once mastered, would provide maximum life-long utility. He concluded that his list lacked sufficient reliability; this encouraged other researchers without discouraging users. Ayres's was seized upon as a source of words for the teaching of spelling and remained popular even after the appearance of Ernest Horn's more definitive word count in 1926. A state speller was published in California in 1914 based upon four small studies done there of word frequency in social and business correspondence (Nicholson, 1914). The flurry of work in reconstructing spelling vocabularies was indicative of dissatisfaction with spelling achievement and the unpopularity of spelling among teachers and students. It was also a reflection of the appeal of "scientific" methods as applied to making the schools more efficient and their curricula more responsive to pupil needs and social utility.

Thorndike's interest in word counts, and his sharpening sense of their uses, was made more personal by the unhappy experiences of his son Edward with a curriculum requiring the daily practice of such words as *marasmus* and *quassia*. A speller of Thorndike's devising, *Growth in Spelling*, appeared in 1929. His early work on word counts was also applied to the reading tests that he authored before the appearance, in 1921, of *The Teacher's Word Book*.³

2.2 *The Teacher's Word Book*

In his historical and bibliographical survey of readability measures, Klare (1963) writes, "The work on word counts set the stage for what was easily the most important occurrence in the period just prior to the development of [readability] formulas, the publication in 1921 of *The Teacher's Word Book* by E. L. Thorndike" (p. 30). This historical judgment, widely expressed, was shared by his contemporaries—who imitated, extended, criticized, and used *TTWB* extensively (F. D. Curtis, 1938, p. 6).

One measure of the influence of *TTWB* has been its use as a research tool in psychology, education, and communications from 1921 to the present. The following investigations suggest the range of its possible

³ *The Teacher's Word Book* will hereafter be referred to as *TTWB* if the reference is general and as *TTWB-1* for the 1921 edition, *ATWB-2* for the 1931 edition, and *TTWB-3* for the 1944 Thorndike-Lorge edition if a specific citation is required.

use. Kennon (1926) employed *TTWB*-1 to construct a vocabulary test of literary terms in a study of selected characteristics of English teachers. Danzger and Halpern (1973) used *TTWB*-3 in selecting words for an investigation of correlations of stuttering behavior and four word traits; they found stuttering ("nonfluencies") positively correlated with longer words and with low-frequency words. Analogous applications number in the uncounted thousands and continue into the present—more than a half-century after the publication of *TTWB*-1 and more than three decades after *TTWB*-3, now the most often cited of the three editions. A frequency count was made of the citations of Thorndike in the 1973 and 1974 volumes of *Social Sciences Citation Index*, a new reference work that employs computer technology to locate and arrange all citations made in the articles, reviews, and notes of a large, international sample of social science journals. The count in these two volumes totaled 234—the vast majority being of *TTWB*, a few others of the Thorndike dictionaries, Lorge and Thorndike's *Semantic Count* (Thorndike & Lorge, 1937), and *The Teaching of English Suffixes* (1941). The combined count of citations of other prominent compilers of word lists and architects of readability formulas—Horn (1926), Buckingham and Dolch (1936), Dolch (1936), Rinsland (1945), and Flesch—was roughly 53; references to Rinsland and Flesch accounted for most of these. Today, investigators in reading, language learning, and language behavior studies of all kinds almost routinely employ a word-frequency list, apparently most often *TTWB*, as their source for the selection and arrangement of verbal stimuli (Carroll & White, 1973; Johnson, 1974). Citations of *TTWB* are especially commonplace in the research reported in the *Journal of Experimental Psychology* and *Journal of Verbal Learning and Verbal Behavior*.

A second measure of the influence of *TTWB* is the impetus it gave the creation of better frequency lists, or those tapping other sources and investigating different patterns of word usage. In an extended description of *TTWB*-1, Thorndike (1921) himself noted several of its limitations. One was the probability that the reading material in which these word frequencies were counted *overvalued* (as did the schools) the languages of literature and morals, as compared with the languages of science, technology, politics, and business (p. 354). Thorndike wrote

that one chief service of the *Teacher's Word Book* will be to aid in the production of some much better list, from wider counts, to replace it. The entire procedure in counting, entering and crediting words is made very much easier, once we have an approximately correct list for use in recording the counts economically. (p. 370)

More than a hundred word lists based on frequency counts were subsequently published (Johnson, 1974). Despite the invitation of the maker of *TTWB*-1 that bigger and better lists be created to make his own obsolete, whatever improvements were made in comprehensiveness of

sources and increases in available words were made through *ATWB*-2 and *TTWB*-3. Not until computers were available to handle millions of words did new general lists appear (Kučera & Francis, 1967; Carroll et al., 1971). Instead, the majority of the later lists were local or otherwise transitory, unimportant except perhaps in the not inconsequential arena of raising awareness of vocabulary issues in pedagogy in the numerous school communities where list-makers were doing their counts, often with teacher and student help. A few lists were quite important and used in various ways by educators, schoolbook authors, and Thorndike himself in each new *Word Book*. Brief discussions of the chief of these should clarify both the character of Thorndike's frequency counts and what other researchers still thought merited being done—if the schools were to be equipped to make better decisions about what they would teach.

2.3 *The Question of Children's Vocabularies*

Experienced workers with children were quick to note that words on the same level of frequency of use in *TTWB* might pose very different degrees of conceptual difficulty for children and adults (G. M. Whipple, 1925; J. C. Dewey, 1931; E. Horn, 1937; I. Morris, 1947). Each of the following pairs of words is composed of words having the same importance in terms of *TTWB* frequency: 2a,⁴ *inside* and *err*; 2b, *ate* and *elate*; 3b, *dirty* and *deem*; 5a, *hers* and *withal*. Given the tendency for the users of frequency word lists to slip from the concept of relative frequency or "importance" to that of relative "difficulty" of words, it was important to have reliable counts of the words that children used; it would then be a simpler matter to calculate the degree of overlap in adult and children's vocabularies for use in grading books. An example of what might be done was Jones's (1915) frequency count of the words written by children in school themes.

Perhaps it was because of the greater receptivity to vocabulary issues found among primary-grade teachers and the continuing disagreements about methods of teaching beginning reading that the more significant early counts of children's words came with very young schoolchildren. The continuing interest in child study and the emergence of child psychology also contributed, as was shown in the sponsorship by the Child Study Committee of the International Kindergarten Union of a study reported in 1928. Under the direction of Madeline (Mrs. Ernest) Horn, kindergarten teachers recorded verbatim the words spoken in the classroom by their pupils and when responding to pictures; records were also kept of words they spoke at home. The "Kindergarten List" repeatedly

⁴2a means that the word falls in the first half of the second thousand in frequency, 2b in the second half of the second thousand, 3a in the first half of the third thousand, and so forth.

found its way into composite lists and was used to evaluate the vocabularies of published primers.

Two examples of word counts of children's writing vocabularies done by Grace McKee and Henry Rinsland will illustrate other dimensions of the activities of gathering, interpreting, correcting, and collating vocabulary data that marked the period 1920 to 1940. Grace McKee (1924), a student of Ernest Horn's, reviewed the previous research on words used in school themes and found it wanting, since it generated very few words that were not on one or both of the criterion word counts: *TTWB-1* and Ernest Horn's still unpublished composite list. Hypothesizing that the ordinary system of assigning theme topics constricts children's language, she sought to provoke freer word usage—and succeeded. The pedagogical implications of this word study were among the clearest offered to the educational world. McKee also kept alive a controversy about the extent to which pupil's actual written work should determine the spelling curriculum. Various critics of studies of children's writing contended that students are hampered in written expression by their inability to spell words; this is such a pervasive problem that it badly biases the research and renders students' writing unsatisfactory as the basis for selecting spelling words. W. H. Coleman (1931) disputed this contention, primarily on these grounds: First, it assumes the existence of a "spelling conscience" in a child that causes him to avoid humiliation by avoiding words he cannot spell—a dubious adult assumption; second, the child probably seldom senses the possibility of substituting a known for an unknown word; and, last, "in the majority of cases his fund of synonyms is so limited that he would have difficulty in finding an adequate substitute if he wished to use one" (pp. 27-28). This debate about spelling words was, of course, a manifestation of a deeper and often silent disagreement about the extent to which the child's needs, wants, and interests can or should constitute a school curriculum.

In an attempt to provide as comprehensive a count of children's writing words as Thorndike's was of reading materials (especially *ATWB-2*) and Ernest Horn's was of adult writings, Henry Rinsland published *A Basic Vocabulary of Elementary School Children* in 1945. Like some of Thorndike's counts in progress during the 1930s, this national study had Works Projects Administration funding. Over 200,000 individuals' papers, of all kinds of uncorrected writings, were collected in 1937 representing the writing of about 1% of the elementary-school population (Grades 1-8) of the United States. Like those other studies which, in some ways, he was following, Rinsland's was a massive counting job: Over 6 million running words were examined. In comparison, *TTWB-1* came from a frequency count of under 5 million words of reading matter, *ATWB-2* from twice that number. The more relevant comparison is with Horn's study of adult writing (1926), based upon slightly over 5 million

words; considering, especially, that children write so much less than do adults, Rinsland's was an impressively large sample of children's writings. It is important also in being a graded list. These young writers collectively used 25,632 different words; if a word occurred 3 or more times in any one grade (14,571 satisfied this criterion), it was included in Rinsland's list—arranged alphabetically with frequency shown for each grade. From his list one may observe (p. 371) that *new* is a word of high frequency and of universal use in the elementary grades, but *newer* is infrequently written before Grade 7; *add* (p. 28) is written some in every grade, especially the third and seventh, but *addend* is not even in this "basic vocabulary," since children evidently do not need or choose to write it.

2.4 *Technical Vocabularies and the Textbook*

"Technical words constitute an important problem in teaching the social studies," wrote Ernest Horn (1937) in a volume of the Report of the Commission on the Social Studies. "Words found in the textbooks of a given field but not in the Thorndike list are ordinarily classified as special or technical words" (p. 165). How did it happen that low-frequency words, that is, words not appearing at all in *TTWB-1*, had come to be equated first with "difficulty" and then with "technicality"? Thorndike had written of the relative frequency of words in terms of relative *importance*—meaning their utility. He had even qualified that relationship a little: "This list is not a perfect measure of the importance of words, for . . . a word may be very important for a pupil or graduate to know and yet not figure largely in the world's reading" (*TTWB-1*, pp. iii, iv). Yet Thorndike himself helped blur the distinction. In both *TTWB-1* and *ATWB-2* he stated to the potential user, the teacher, that "knowledge both of the importance of the words and of their difficulty" could be gotten from the contents of *TTWB*. In a study done with Symonds and reported in 1925, Thorndike again remarked that difficulty and frequency are not synonymous, as shown by the less than perfect correlations obtained when schoolchildren were tested with words on a Thorndike vocabulary test (based upon *TTWB-1*). He also noted that a word may be difficult in one context, as in a test or list, but not in a more typical reading situation (Thorndike & Symonds, 1925, p. 440).

A pattern of research developed that selected words from the several frequency levels of *TTWB* to test students' word knowledge—some words selected from the first thousand in frequency, some from the second thousand, and so on, with whatever variation the experimenter had designed (Yoakam, 1927; Patty & Painter, 1931; Gates, Bond, & Russell, 1938). Other researchers selected words of high frequency from *TTWB* in order to rewrite material; equating high frequency with ease of reading, they mistakenly called their revision "simplified." They then tested children and compared their comprehension of the original with the sim-

plified version. If they found little comprehension improvement with the revision, or even a decrease in comprehension, they concluded that *words* are less important in determining readability than heretofore believed (Nolte, 1936; Serra, 1954) and that readability must depend heavily on such factors as syntax.

Powers (1925, 1926), in his word counts in science, also displayed this tendency to equate frequency/importance with difficulty. He cited Thorndike and Symonds' (1925) finding of a positive correlation between infrequency and difficulty, and their observation that ninth graders were estimated to be able to recognize two-thirds of the 10,000 words in *TTWB*. Then Powers (1925) concluded that "words of such infrequency of use that they do not occur in the 10,000 most important words may be presumed to occasion considerable difficulty to ninth grade readers [and] a relative measure of the difficulty of text-books is, therefore, the number of words that are used in them which are not found in the *Teachers Word Book*" (p. 368). A similar line of reasoning was taken by the social studies director of the Brookline, Massachusetts, schools. He had teachers keep records of the words used in teaching social studies that caused difficulty; it was found that 48% of these words were not in *TTWB*. By projection he concluded, therefore, that the vocabulary load of junior-high-school social studies was so great as to be "beyond the comprehension of at least 50 percent of our pupils" (Kepner, 1929, p. 32).

Science education was a predictable place in which to equate low-frequency words with technical words—using *TTWB* to identify the former (Rader, 1922). Francis Day Curtis was the most often cited, and probably the most vigorous, early disseminator of vocabulary research knowledge in science education (Curtis, 1926a, 1931, 1939). His students at Michigan were busied with examinations of the language of the science curriculum, 40 of them completing various vocabulary studies of textbooks during 1930–1931 (F. D. Curtis, 1938, p. 21); George Mallinson (PhD, Michigan, 1947) inherited the cudgels in the fight to lighten the excessive vocabulary burden of the science curriculum (Mallinson, 1952). Himself a Teachers College graduate student when *TTWB* was a widely employed tool in graduate research, Curtis in his dissertation (1924) analyzed 630 newspaper science articles; one item of analysis was determining the placement of their scientific terms within *TTWB*, if they were included. Curtis's next study (1926b) compared the proportion of scientific terms to the total vocabulary of these science articles and that total vocabulary to *TTWB*; he concluded that teachers should take steps to provide spelling drill and word building—of the *scientific connotation* of many common words as well as of the technical terminology of science.

The fact that scientific terms overburdened the textbooks, that the terms were ill-defined, and that they were given insufficient repetition to be learned was reported in 1925 by S. R. Powers, professor of natural

science at Teachers College. His results led him to construct a special word count of scientific terms for the use of science teachers; from a running count of over 2 million words from textbooks, popular scientific books and magazines, and miscellaneous science writing, he produced a list of 1,823 words of highest frequency and range of use (Powers, 1926). His was admittedly modeled after *TTWB* and intended as a pedagogical supplement to the Thorndike list.

Even in the school subjects not thought to possess much of a technical vocabulary, word counters—armed with *TTWB* or the supplementary vocabulary lists which had begun to proliferate—were reporting what they typically described as an overdose of textbook language of conceptual difficulty (Ward & Stevenson, 1925; Ayer, 1926; Mathews, 1926; Moser, 1927; Pressey, 1931). The fact that the special vocabulary of the several content fields corresponded to the situation in science textbooks was repeatedly noted by researchers, educators, and authors of educational psychology texts for teachers (Buswell & John, 1931; Cole, 1934; F. D. Curtis, 1938; Reed, 1938). A 1930 study (by J. C. Bedillon) of elementary-school history books stated that pupils were being confronted with the equivalent of learning a foreign language vocabulary of 3,000 words (in Seegers, 1939, p. 322; see also Brown, 1931). Another study reported a commonly encountered thoughtlessness on the part of textbook authors: An English textbook used most in the senior year was found "easier" in regard to vocabulary than was a freshman text (Patty & Painter, 1931, p. 31). New words were discovered to be introduced almost absentmindedly and difficult words, once used, might never reappear in a text. Teachers were advised of their duty to provide for the repetition or opportunity for practice upon which mastery ordinarily depended—this to compensate for the textbooks' failure.

In the word study of Bertha Lively and Sidney Pressey one can see the coming together of pedagogical need and objective science. The need was expressed in teachers' complaints about the difficulty of science books used in the junior high schools; the science was expressed in the tool of *TTWB*, of which Lively and Pressey (1923) wrote, "The 'Word Book' has opened up a whole new field for investigation" (p. 396). These researchers devised a strategy that they apparently hoped would yield a more sophisticated, composite measure of the relative difficulty of a piece of reading material than raw counts of the number of words inside or outside of *TTWB*. They hypothesized that three questions should be asked about the words constituting a sample of reading matter. First, how many different words are found? This yields a measure they labeled "range." The range for the Jones second reader was 371 or 359 per thousand depending upon which of two samples was examined; for Muzzy's *American History*, the range was 533 or 536 per thousand words. Second, where do these words fall in the successive levels of *TTWB*, that is, what are

relative statuses insofar as word-frequency data are concerned? A word falling in the first thousand of *TTWB*-1 got an index number of 10, a word in the tenth thousand got the number of 1; then a median index number was computed. Third, what is the number of "zero-value words," that is, words not found in *TTWB*? *This was considered the measure of technical difficulty or the extent of the technical vocabulary.* Figure 1 illustrates the data calculated from two 1,000-word samples: each of six science texts, a history text, a newspaper front page, Thackeray's *Vanity Fair*, Stevenson's *Kidnapped*, and two readers.

A calculation based on the index numbers and zero-value index gave a "Weighted Median Index Number." Lively and Pressey concluded that it probably best expressed what they called "vocabulary burden"—a term that became a cliché; variously used in the literature, it seems to have been more often thought of in pedagogical circles as the number of words in a book or the proportion of new words to old. However important the Lively and Pressey study was in coining an appealing popular term, it also has theoretical importance as an early example of an attempt to calculate and weight several components that might determine readability. Note that each component in this "formula" concerned vocabulary. Thus, it is unlike those later readability formulas that took measures of such other factors as average sentence length, parts of speech, or personal pronouns (Chall, 1958a, pp. 42-43; Klare, 1963).

Luella Cole is the final example to be offered here of a vocabulary researcher who made important contributions to the movement of vocabulary research into the schools. By her own account (Pressey, 1934, p. 155), hers was also a response to the experience of scholastic frustration—in this case from supervising remedial programs for college students. Cole's analysis convinced her that a chronic inability to master textbooks, especially in the social sciences, rather than a lack of general reading ability, was a significant factor in college difficulties and that a vocabulary-building program was in order. In 1924 she reported the results of a study of 100 textbooks in several school subjects (Pressey, 1924a). Her aim was to construct a list of the essential technical vocabulary with which the teacher in each field could begin teaching what she called "assimilative reading" (Pressey, 1924b).

The poor achievement being reported in school arithmetic and the manifest dislike of the subject had, as in the case of spelling, already raised some interest before Cole published her arithmetic vocabulary as one of 19 special vocabularies (Pressey, 1924b). Some experts had even proposed that "book arithmetic" be postponed "until the difficulties of primary reading are well out of the way" (LaRue, 1924). On the need to know the role in arithmetic of reading difficulties there was wide agreement (Monmouth, Asbell, Jenkins, & Pieters, 1924; Stevenson, 1925; Garrison & Garrison, 1929; Pressey & Elam, 1932). The belief in mean-

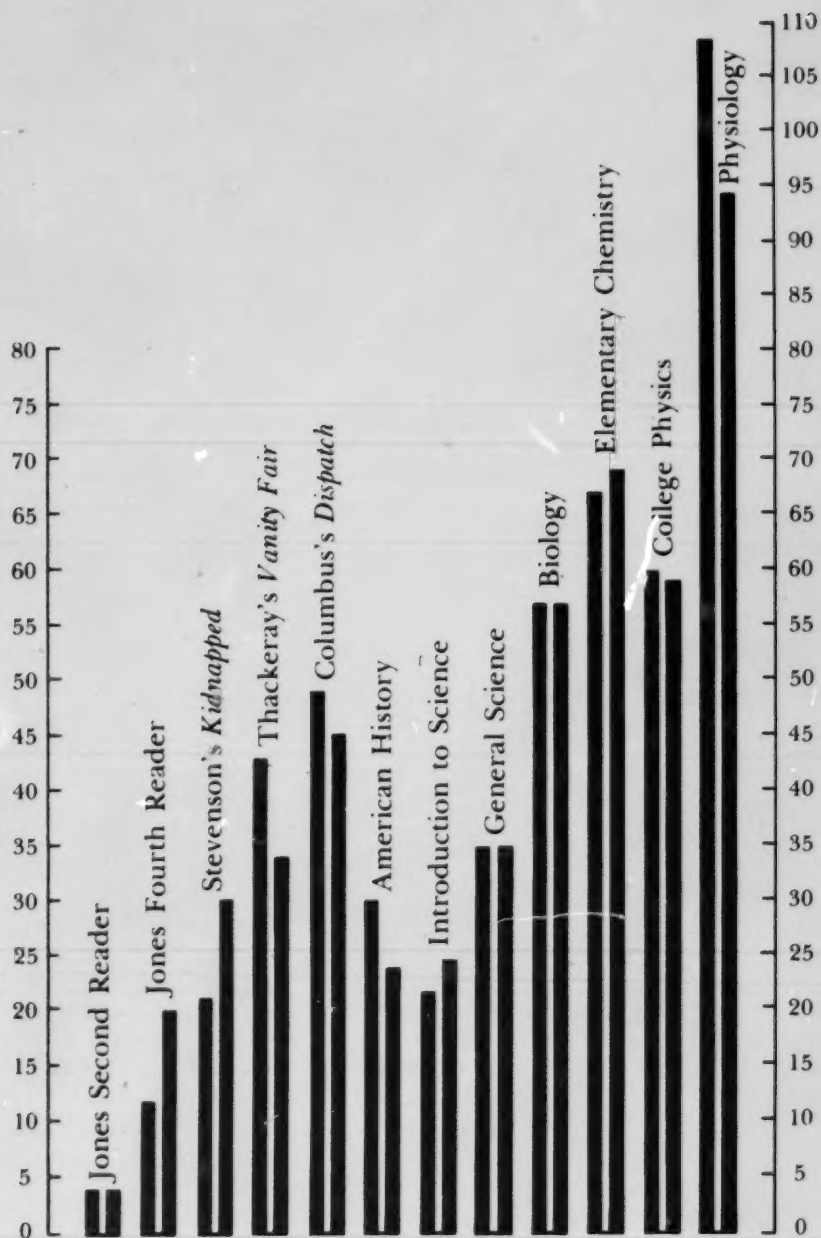


FIGURE 1. Number of "zero-value words" of reading material of different types (1,000-word samples; Lively & Pressey, 1923).

ingful reading applied also to arithmetic reform, and it became axiomatic—with the help of the *Thorndike Arithmetics*—that it is better in children's arithmetic to read of "3 chocolate bars that sell for 10 cents" than of "two liters of petrol that cost 9 francs" (Reed, 1938, p. 314). That vocabulary decisions are not solely a matter of simple word frequency is illustrated by Cronbach's (1955) example: "The child can learn to add just as well if he never hears the word *addend*," but "the words *numerator* and *denominator* are indispensable for mastering fractions" (p. 82). Such a position was Thorndike's also.

3. CONCLUSIONS DRAWN FROM FREQUENCY RESEARCH AND THEIR IMPLICATIONS FOR EDUCATION

It may be true, as Harry Levin (1966) maintains, that the mass of reading research has produced few conclusions for the edification of practice (p. 138). Whether or not the following list of agreements among the students of word frequency qualifies as a set of "trustworthy generalizations" is not known; they do appear to represent a consensus of informed opinion. This consensus was largely achieved within the first two decades after 1920. Furthermore, school practice, in the language areas in particular, did come to show some consistency with the implications of these facts.

3.1 *Overlapping Vocabularies and Curriculum Integration*

An early goal of vocabulary researchers was to establish the relationship between the vocabularies of hearing (recognition), speech (oral), reading, and writing. Word-frequency counts of each appeared early, and there was considerable cross-fertilization of methods. In 1926 Ernest Horn had noted that studies of the speaking vocabulary were still too few; that remained the case until recorders were available to supplement or supplant the stenographers. As word counts multiplied and as research and testing provided more data on growth in each vocabulary from year to year, estimates of the initial disparities were reduced. While there was growth in each vocabulary, there was also a reduction in the differences between them. The estimate of a once large gap between the recognition and writing vocabularies, for example, was lessened, as was the initially more modest gap between reading and writing (Washburne, 1923; Yoakam, 1950, p. 126). This was especially the case for the words of highest frequency of use. A study reported in 1935 obtained stenographic records of fourth graders' conversations and found that 89% of the words used fell in the first thousand of *TTWB* (Durrell & Sullivan, 1938, p. 140). By the seventh grade, M. K. Smith found that students could read most of the words they could recognize when heard (Seegers, 1949, pp. 183–184). Ernest Horn (1926) had predicted that once sufficient sampling was done

of the reading, writing, and speaking vocabularies, a much greater overlap would be found than some had assumed (p. 188). Horn would not then have been surprised at the recent study (Chester & Otto, 1974) that found that the words children use in speech and the words in their readers are more nearly alike than the critics of today's reading books claim. One explanation is that children parrot the words they read in repetitive basal series. Another explanation, and it would be Horn's, is that the several vocabularies naturally duplicate one another: People write and read the words they hear themselves and others speak, and people's speech includes what they have read and written. The third explanation is historical: Studies of children's oral language have been used in the construction of school reading materials, especially in the primary grades.

The word counts of each one of the several vocabularies have had some obvious independent or singular effects upon some relevant curriculum areas; the reading curriculum was affected more by word studies of reading vocabularies than by the speech or writing vocabulary research. But influence does not stay neatly tracked and confined to its obvious home. The studies of writing vocabularies did greatly alter the character of published spelling books, city and state spelling lists, and courses of study (Coleman, 1931; Wise, 1934, 1935; Horn, 1941; Dickerson, 1950; Greene et al., 1953; Hildreth, 1955; Dale, 1956, p. 114). Insofar as vocabulary research was concerned, these changes were effected in a uniform and drastic reduction in the number of words included in the spelling curriculum, in a general removal of the less common words, and in a less-than-standardized progression from common to less common words. Spellers, however, were also changed by studies of children's speech vocabularies; data based on over 600,000 running words were available by 1926 (Horn, 1926, p. 188). Betts (1949b, p. iv) states that the use of oral vocabulary data in selecting words for inclusion in primary-grade spelling books helps account for the fact that the word contents of different published spellers agree more in the primary spelling curriculum than they do later on. Ayer (1945) found some evidence to suggest that high school spelling books were using *TTWB*, a reading vocabulary study, more systematically than they used Horn's writing vocabulary list. This latter finding, if substantiated by more direct evidence, should not surprise anyone, however, for the publishers of textbooks appear competitively disposed to select their content or adjust their words from the best known, most prestigious list; what mattered was that their source was scientific.

Moreover, some of the better known word lists, especially for elementary-school use, were themselves composite lists drawn from word-frequency data on writing, speech, and reading. Buckingham and Dolch compiled their aptly named *A Combined Word List* (1936) from 10 sources—covering writing, speaking, and reading vocabulary studies—plus their own "free-association" study of writing vocabularies. Another

combined list, representing reading and speech sources, is Arthur Gates's *A Reading Vocabulary for the Primary Grades* (1926/1935). The earlier edition was based on the highest 2,500 words from *TTWB*, a study of the vocabulary used in 10 first readers, and a study of young children's speaking vocabularies. The revision took advantage of intervening studies of children's reading materials but retained the speech vocabulary component. As Gates explained it:

Since the words in this list are used in speech, are found in primary readers and in selected primary literature, are related to interesting and useful features of children's lives, and are also words most frequently encountered in later reading, they should be especially worth learning, easy to learn, and once learned are likely to be retained. . . . By using the same words as far as is possible in reading, writing, and spelling, it should be possible to develop language ability with greater economy of time and effort. (pp. 2-3)

Gates is universally described as having contributed, in his *Reading Vocabulary*, what became almost an indispensable list for the purposes of rating or rewriting the content of primary-grade school readers.

A conclusion of research was that as children mature and are educated their several vocabularies converge to create a considerable overlap. This means, as Gates suggested in the last sentence above, that words useful for one language activity in the classroom have a high probability of functioning well and being meaningful in another. This supports a breaking down of compartmentalization in the curriculum. As Gertrude Hildreth (1948a) noted, in a comparison of the Dale (recognition), Dolch, and Rinsland (writing) vocabularies, "There is an increasing tendency in school practice to combine the reading and writing vocabulary for easier and more permanent word mastery. The combined word list lends itself well to unified language teaching in the lower grades" (p. 45). The "language arts" approach grew to popularity in the 1940s and 1950s. The principles of an integrated approach (and the fusion of formerly separated activities was what integration meant in the 1930s and 1940s) were described by a prominent reading expert and student of readability and vocabulary (Yoakam, 1950):

As the child grows, listening, speaking, reading, and writing vocabularies should grow simultaneously, although not equally, at any given time, the reading vocabulary at the beginning should be based on the listening and speaking vocabularies, and the writing vocabulary should emerge from listening, speaking, and reading. . . . All during the elementary grades listening, speaking, reading, and writing should form a group of language experiences in which listening and reading give the child experience with language, speaking clarifies meaning, writing gives more exact expression to what the child thinks, and in which listening, reading, thinking, and discussion naturally leads to the desire to express ideas in writing. (pp. 18-19)

The language arts approach became the orthodox view in elementary schools. The effort to achieve similar ends was extended to the secondary

schools; as advertisements suggest, there was some success in the junior high schools despite the much-commented-on resistance of many English teachers to curricular innovation. It was one thing to accept the assumption behind the *From Thoughts to Words* lesson supplements published in the 1960s by the National Council of Teachers of English, "that young children's listening and speaking vocabularies highly exceed their writing vocabularies" and that primary teachers ought to be interested in word activities. It was something else for status-conscious teachers of high-school content subjects, where many English teachers placed themselves, to be persuaded that they should each be teachers of reading, writing, and speaking, as well.

3.2 *Overlapping Adult and Child Vocabularies and ESL Instruction*

As frequency counts multiplied following publication of *TTWB* in 1921, comparisons of their findings were promptly made to test the presumption that children have their own vocabularies. In 1926, Ernest Horn surveyed the accumulating studies of the spoken vocabulary of children but found the data on adult speech too meager. But, he concluded, "a tentative comparison of the spoken vocabulary of children and the written vocabulary of adults shows a very large overlap" (p. 188). As studies of children's writings grew in number, and in extensiveness with the publication in 1945 of Rinsland's findings, Horn (1950) examined the results. "There is a significantly large overlap between the words frequently written by children and those frequently written by adults. This overlap is especially marked between the words used in children's letters and those used in the personal letters of adults" (p. 1251).

The least agreement between adult and children's vocabularies is in the nouns they use, the names for "things." Fries (1950, pp. 75-76) provides some examples of quite disparate frequency placements when some *ATWB-2* words are compared with those of the Kindergarten Union (KU) study of the 2,596 words children apparently find most comfortable to use: *Paste* appears in the fifth thousand in *ATWB-2*, in the first five hundred in KU; *sticky* in the *ATWB-2* twelfth thousand, but in the second thousand of KU; *bath* in the third thousand of *ATWB-2* and in the second five hundred of the KU list. Similar examples could be offered, including words standing for the more important experiences in the child's world, words like *ice cream*, which does not appear in *ATWB-2*, and *tricycle*, found in the fifteenth thousand. Dale (1931a) listed the 769 words from the KU list also found in the first thousand words of *TTWB-1*, and the 230 words of *TTWB*'s first thousand not found on the KU list—words like *although*, *coast*, *sight*, *enjoy*, *voice*, and *however*.

For reasons of important if not extensive differences, it became obvious that lists of adults' words of highest frequency had to be augmented by "children's words" for instructional purposes. This was natural for

nominally child-centered schools that promoted a pedagogy of meaningfulness and experience; at the least the teacher, the curriculum, and the book were expected to "start from where the child is." Conversely, it was inappropriate to give adults learning a new language a child's primer; aside from inappropriate conceptual content, certain high-utility words for adults would escape being encountered early enough. In outlining the possible uses of *TTWB*-1, Thorndike (1921) suggested:

By the elimination of certain specially childish or "literary" words from the first 500 of our list and the addition of certain words of special importance to the newcomer to America . . . we shall have a basic list of great value in teaching foreign adults to read English. A tentative first 500 for foreigners may be formed simply by omitting *ball*, *pretty*, and *soldier* from the list on page 365, and adding *danger*, *poison*, *cent*, and *dollar*. A second 500 can be formed in much the same way. The use of two such lists, with any additions necessary to arouse interest and to meet local needs, may be expected to improve and facilitate the teaching of English to foreigners above the age of twelve. (p. 364)

In 1930 the New York State list was published as a beginner's reading vocabulary for foreign-born adults, especially those illiterate in any language. *TTWB* and the Gates list were among its components, along with other words selected as essential to adult activities (Cornell, 1930). Today one reads recommendations to English teachers that a minimal list based on high-frequency words plus a small supplement be taught in high school remedial and adult illiteracy programs (Johns, 1972). A distinction between the initial language learning of a child and the deliberate instruction in a second language, as in the case here, was put cogently by a participant in the large, recently computerized study mentioned earlier (Kučera & Francis, 1967): The child "learning his native language is, as it were, bathed in a continuous stream of language. . . . The person being taught a language is more or less systematically exposed to a more or less carefully selected sample of the language" (p. 134). Foreign-language teachers have been, therefore, consistently more sensitive to vocabulary than the average teacher.

The teaching of English as a second language, either to adults or to bilingual schoolchildren, merits fuller discussion than it can receive here; it is only fair to the vocabularists, however, to indicate that there is ample evidence of their various researches' having entered into the linguistic and pedagogical discussions of an international community of interest that is English-language educators (Coleman, 1938; Fries, 1950). Thorndike's place in that community is shown in three primary ways. One was the use of *TTWB* in the preparation of textbooks for adults, most notably in the books of Lawrence Faucett and Michael West (West, 1973), and, through the Rejall reading vocabularies, the Americanization of textbooks that reportedly used the Rejall vocabulary (Rejall, 1926; Tharp, 1931; Faucett

& Maki, 1932; Thorndike & Lorge, 1944/1972, p. xiii). Thorndike's effect was also felt through the use of *TTWB* in the readability formulas applied to measure the difficulty of English-language reading materials for foreigners. Third was Thorndike's participation in devising the *Interim Report on Vocabulary Selection* (Faucett, Palmer, Fife, West, & Thorndike, 1936), which was of considerable interest to teachers of the several modern languages (Coleman, 1938). This list resulted from a conference in 1934 which convened leading American and British language specialists, especially in education.

The political and social factors that created an international colleague group are worth mentioning. An early center for the utilization and promotion of vocabulary research was the Institute for Research in English Teaching founded in Japan in 1923 by an Englishman, Harold E. Palmer, his countrymen, and the Americans who were found so ubiquitously in Japanese middle schools and universities. Japanese government officials and business leaders had been promoting English-language competence since the decisions to open Japan to the West in the 19th century; by virtue of compulsory English instruction, Japan was virtually bilingual in reading before the Second World War (Coleman, 1938; Crocker, 1945). The Institute's *Bulletin* (BIRET) reported Palmer's and others' efforts at distilling the most useful words of high frequency for teachers avid for simple materials like the cheap readers found in American variety stores—"The Mailman," "My Pets," "The Locomotive."

The presence of a large Spanish-speaking population in the American Southwest was reportedly another early stimulus of research directed to compiling a basic or fundamental English word list (Kavetsky & Morrison, 1960, p. 478). In the mid-1930s Tireman and his associates at the San Jose Experimental School of the University of New Mexico constructed a 500-word English vocabulary from *TTWB*, Gates, KU, a study of the first readers used at the school, and a few local words; similar efforts went on elsewhere (Tireman, Dixon, & Cornelius, 1935; Tireman, 1941). This was an era when the United States was still assimilating millions of the foreign-born and the native-born American speakers of other languages, when Canada and Australia were receiving large immigrations, when Great Britain was imposing her language upon a vast colonial empire in Africa and Asia, and when the United States was being the "Good Neighbor" of Latin America. The radio and phonograph crossed political borders with language, and an interest in finding or creating some one "international language" was renewed. Esperanto was one candidate. Another was "Basic English," which stimulated enough books and articles in journals that a separate entry appeared in the *Education Index* in 1932; Ogden's book, an 850-word vocabulary, was published in 1934.

3.3 *The Skewed Distribution Curve of Word Frequencies and Practical Instruction*

Vocabulary research confounded those who might have expected that word frequencies would distribute themselves along the omnipresent "normal curve": a few words seldom used, a few words regularly used, and the great number of useful middle-frequency words clustered comfortably together. Had the facts about vocabulary usage supported the bell-shaped, normal-curve theorem, the message to the schools—and to the authors and editors of textbooks—would have been essentially that they keep on doing what they had been doing: introducing a great number of words in each school subject and grade because there are so many serviceable words to be learned and so little time.

Instead, *what was found was a drastically skewed distribution, and the implications for vocabulary teaching and learning were quite different from existing practice.* The vocabulary curve has been aptly described as being "like a ski jump: a very few very-high-frequency words, a small number of medium-frequency words, and a very large number of very-low-frequency words" (Twaddle, 1966, p. 76). The results Leonard Ayres had reported in 1913, after a count of words used in adult writing, were duplicated by all later studies of word usage, spoken as well as written. Thorndike's psychological system and his extensive personal experience with word counting led him to conclude:

There surely are forces acting to cause frequent use of the same word. Certain persons, things, events, and relations occur often in thought and communication. Other things being equal, the use of one symbol for one person, thing, event or relation saves much time and effort for the speaker-writer, and still more for the hearer-reader. (Thorndike, 1937a, p. 406)

Probably to his own amazement, Ayres (1913a, 1913b) had found that just nine words contributed one-fourth of the total number of words written; 50 words, and their repetitions, made up one-half of all the words used in writing and, except for the word *very*, all are monosyllabic words (Ayres, 1913a, p. 81). Since the 300 most common words constitute 75% and the 1,000 most common constitute 90% of the words we write, it quickly becomes clear that the population of words whose mastery teachers should seek is much smaller than formerly imagined. Conversely, it is more evident than it was before that mastery of the very-high-frequency words is essential for effective linguistic functioning—"mastery" entailing instant visual and auditory word recognition, quick and correct pronunciation, habituated proper spelling, and the ability to choose among their several meanings upon encountering words in common usage.

Subsequent frequency studies confirmed Ayres's general findings. Ernest Horn (1926), comparing his own studies of writing vocabulary with

the words of reading in *TTWB* and the few extant studies of the words of speech, concluded that a 500-word vocabulary would be a highly efficient tool for teaching English to the foreign-born; selecting the 500 words common to all three lists would provide learners with the words which, with their repetitions, make up three-fourths to four-fifths of all the running words used in the language (p. 197).

Because Rinsland kept data on a word's frequency in writing in each elementary-school grade, it was possible to check progression from grade to grade in the variety of words employed, as total vocabulary (number of different words) grows. Thus, the number of different words used by pupils increased from 5,099 in Grade 1 to 17,930 in Grade 8. In Grade 1 the first 100 words in frequency constituted 63% of the total number of running words; in Grade 8 it had decreased to 57%. Comparisons of the proportions of total running words that are contributed by the most common 100, 500, 1,000 words, and so on, are remarkably similar from Grade 1 through Grade 5; thereafter the dependence on the most common words diminishes a little. This difference is probably a reflection of several factors, including (a) more exposure in the upper grades to the technical vocabularies of the content subjects and a sufficient mastery of some of these words so as to enable students to use them in writing; (b) the influence of word-building exercises to further sufficient knowledge of word roots and derivatives; and (c) the likelihood of greater facility in dictionary use and reference reading, particularly for use in school writing assignments. Nonetheless, the fact remains that, even in Grade 8, the most common 2,000 words, and their repetitions, account for 90% of the more than one million running words of eighth-grade writing; the corresponding percentage in Grade 1 was 98% (Rinsland, 1945, pp. 10-12). As Rinsland shows, a small number of words do yeoman duty. In virtually all counts, the 500 most common words do 65% or more of the work (Lorge, 1944, p. 549).

The most recent frequency data confirm the earlier conclusions about the general outline of word distribution; the asymmetry of word usage in favor of a very small part of the lexicon receiving endless repetition is a confirmed fact about the present users of the language, as it was about those in the past. The American Heritage Intermediate Corpus, computing frequencies in over five million running words of the variety of reading material that children encounter in schools, determined that the most common 1,000 words account for 74% of all running words; the first 5,000 words (equivalent to half the *number* of words in *TTWB*-1) account for 89.4% (Carroll et al., 1971, p. xxvii). The computer count of Kučera and Francis (1967) located 50,457 different words in a one million word count. Most of that 50,457-word vocabulary was seldom used: 42,540 appeared 10 or fewer times, and nearly half the vocabulary (i.e., 22,598 words) was found to occur only once in over one million running words.

Given the broad sampling of types of reading materials used to provide words for this count—from detective and romantic fiction through sports news and industrial reports—there is ample confirmation of the finding remarked on by Gray and Leary in 1935, that words having high frequency of occurrence in *TTWB* would have high usage in all types of language materials (p. 275).

The skewness of the distribution of word frequencies meant, both for researchers and the users of research, that a high degree of reliability could be attached *only* to the placements of the very common words. The extension of *TTWB* published in 1931 added five million words to the running count and broadened the coverage to achieve a less literary word base; the extension, however, was thought also to improve the reliability of the frequency placement of the first 10,000 words. By 1950 the consensus seemed to be that "reasonable reliability" marked the Thorndike list of the 10,000 most frequently used, nontechnical words (Greene, 1950, p. 389).

As Thorndike and other word researchers well knew and frequently stated, as one encounters words beyond the most common one thousand or so, placement at one level or another of frequency becomes progressively less reliable as the word becomes more rare; Clarence Fieldstra's comparisons of *TTWB*-1 and *ATWB*-2 revealed some significant shifts (Curtis, 1939, pp. 18-19). In *TTWB*-3, the words marked "AA" were the 1,069 words occurring at least 100 times per one million running words; conversely, there were 11,221 words listed which appeared five or fewer times per million, about half of these appearing once. Placement of such uncommon words as *absentee*, *cavalcade*, *forefoot*, *nauseate*, *sidetrack*, and *Theodosius* is too much a matter of the vagaries of sampling for teacher- or author-users to rely upon extensive lists (Thorndike & Lorge, 1937). This seems a major reason for the considerable variability in content and placement found among the modern spelling books—all claiming to include the 3,500 to 4,000 most common words, as determined by statistical studies (Wise, 1934; Lee & Lee, 1940). This was foolhardy and pretentious, given the clear warning of the author of a major source for such spelling lists, Ernest Horn, that after the first thousand words or so, the reliability of a count drops quickly (1926, p. 20). There was a danger from the same cause for researchers who used vocabulary tests of words drawn from the different levels of frequency for the purpose of estimating the total size of a child's vocabulary (Thorndike & Symonds, 1925).

Even before the publication of *TTWB*, the pedagogical advantages of using frequency lists as the source of word selection had occurred to several educators, particularly in the field of language teaching; Lodge was an example in the classics. Palmer (1917), when still in England, had made the observation that frequency as a criterion for vocabulary selection was merited on the grounds both of *utility* and of *pedagogy*. By the

latter he meant that repetitious use of the most common words conformed to a basic principle of teaching and of learning; the fact that they were the most common words itself aided their being learned (pp. 122-125).

One of the unquestionably significant consequences of the production of the first word-frequency counts was their application, often by teachers and teachers-in-training, in evaluating the appropriateness of the books used in ordinary classrooms (Brown, 1920; Gregory, 1923; Buswell & Judd, 1925; Irmina, 1929). Examples from the findings of two studies reported by Thorndike are reproduced in Tables 2 and 3. Such examinations consistently revealed two facts about the majority of textbooks—facts purportedly less true as time passed and textbook authors and editors yielded somewhat before vocabulary research. One was that most textbooks contained a great many different words; therefore, they were certain to be tapping too many of the less common, less important words (Uhl, 1924, p. 21; Coleman & King, 1941, pp. 528-529). The second fact was that, for learning purposes, textbooks were making inadequate provision for repetition of the less common words, a repetitiveness that if not provided by deliberate decision would not be provided "naturally"; uncommon words are, by definition, not often encountered. Thus, traditional textbooks were imitating life; but in this case, pedagogy should *not* imitate life.

TABLE 2
Words in the First, Second, and Third Readers Which Are
Not in the 10,000 (Thorndike, 1921)

backwoods	bluejay	calmy	cleak
benighted	bluebird	Campbell	climax
baff	blumb	cantaloupe	climber
bagful	Blynkin	Canute	clop
baggy	bobtail	carman	clump
bab	boo	Caroline	cockshell
bannock	Borneo	Carrara	columbine
barbecue	Bose	Caspar	commoner
barrelful	bossy	Charlotte	Conrad
basketful	Bowden	chatterer	Cora
beanshooter	bricklayer	cheep	corncob
bearskin	Brinker	chipmunk	Corrola
beehive	brooklet	chirrup	counterpane
Bert	broomstick	chore	craggy
Bess	Bruckner	christen	crawly
Biddy	burdock	chubby	creepy
biff		chug	crispy
bight		chump	croon
birdling		churnful	crotch
blithesome		citron	crumple

TABLE 3
Words in the First Fifty Pages of a Standard Algebra Which Are
Not in the 10,000 (Thorndike, 1921)

abbreviate	Diophantus	literal	prefix
Ahmes	dissimilar	mathematical	Pythagoras
algebraic	distinctive	mathematician	quadrilateral
algebraist	elementary	minuend	rearrange
arithmetical	et	monomial	redwood
assets	exponent	multiplicand	reintroduce
binomial	facilitate	multiplier	rewrite
bricklayer	formula	ninth	scalespan
casting	formulae	nitrogen	simplify
coefficient	Harriot	numerical	Stifel
complementary	haw	Oughtred	subtrahend
computation	Herigone	parallelogram	supplement
debit	Hindu	parenthesis	supplementary
Demosthenes	ingenuity	Pell	trinomial
denominate	integer	perimeter	Vieta
Descartes	Leibnitz	polynomial	Widmann
digit	likelihood	potentia	workmanlike

The pedagogical implications of the findings on word-frequency distribution were probably as clear as the whole of vocabulary research could offer. One lesson from the work was that the first priority in teaching words should be teaching what Michael West called wide-range words. It was Thorndike's contribution to the methodology of vocabulary counts to add range to frequency as a measure of a word's importance (Fries, 1950, p. 21). While frequency indicates how often a word is used, range refers to the coverage a word receives over a number of sources—how widely it is used. A word found once in 50 sources is a more important word than another mentioned 50 times in one place. A manifestation of the effectiveness of the vocabularists was the growth of a common textbook vocabulary within and across the various school subjects. The early analyses of textbook content frequently commented upon the fact that textbooks had so little common vocabulary, even in those books on the same subject offered for the same grade level. Later comparison found a greater degree of commonality within curriculum areas and between them. A higher proportion of the total vocabulary of schoolbooks (and probably of courses of study) was composed of wide-range words than before.

A second and even more obvious message from vocabulary research came from studies of word frequencies, investigations of pupils' problems in word comprehension, and surveys of vocabulary extent and development: Reduce the vocabulary burden of almost all textbooks. A dramatic example was offered by spelling. The old spelling books had resembled

small dictionaries in their numbers of words; some were reported as containing as many as 15,000 words (La Rue, 1924, p. 177). "It is a frequent defect of the quality of scholarship to crave to fill out a system to symmetry and completeness," Thorndike (1934) observed. "So makers of spelling books used to search the ends of the linguistic earth to find words that sounded alike but were spelled differently, and perpetuated such atrocities upon childhood as *new gnu*, . . . *color culler*, and *fees fease*" (p. 103n). By 1925 a new consensus showed itself in spellers: The total spelling curriculum for Grades 2 through 6 or 8, should contain from three to four thousand words; the wise teacher will attempt to teach only about 400 words per year (Hildreth, 1955, p. 127). The excised words should include all the rare and terrible words—*abattoir*, *acciaccatura*, *Houyhnhnms*—that had no utility, even for the college graduate. Most of the less common words could go also, especially if sensible school dictionaries existed and their effective use was taught. Not long ago, Dolch recommended, in fact, that the common spelling curriculum be no more than the 2,000 most common words and that it be supplemented by the most important words in the content areas and by those servicing individual interests (Lee & Lee, 1950, p. 423; Dolch, 1960, p. 9). A Catholic educator surveyed the spelling scene in 1930 and suggested use of some standard list augmented by words specifically used by Catholic children (Foran, 1930). The opportunities for promoting vocabulary building and for providing for repetition are obvious in this scheme. It also conforms to the ideologies of individualization and the integration of subject matters, ideas that were so much promoted in teachers colleges, in school-district courses of study, in professional conferences, and in the educational journals.

What of the low-frequency words? If they do not recur enough to be learned, should they be abandoned and, along with them, the goal of vocabulary enrichment? Given the demonstrated distribution of words, there is an admittedly difficult learning problem in vocabulary. In *ATWB-2*, Thorndike (1931/1932) acknowledged this, calling attention again to the problem of lack of repetition for the rare words of the language:

This rarity, the degree of which has probably not been realized hitherto, is of obvious importance in education. If we suppose that a person increases his vocabulary by learning words roughly in the order of their frequency of use, his task tends to become easier and easier with successive thousands because of the increasing number of derivatives. But it becomes increasingly harder because of the rarity of reviews. To acquire and keep alive a vocabulary of ten or fifteen thousand words over and above the five thousand commonest is a hard task. (p. 182)

It lessens the difficulty if, among the infrequent, the truly important words are selected for practice. The teacher can assist in providing the

practice; so can textbooks which are more conscious of, and attentive towards, vocabulary in their design. As Faucett and Maki (1932) put it, "Wide range words are determined by the language; narrow range words should be deliberately chosen" (p. 7). Readable dictionaries and carefully considered drills in word attack will also help foster greater word independence. Perhaps the most important "remedy" the schools can offer is to teach for better use of context—although context is of no help if there are too many rare words, as in this example from the Nicholas Pike arithmetic (1788; in Uhl, 1924):

Deduct the tare and trett, divide theuttle by 168, and the quotient will be the cloff, which substract from theuttle, and the remainder will be the neat. (p. 12)

As in all educational work, the school must be pragmatic: It must cease thinking of words either as equivalent in importance or as all to be brought to the same level of mastery. Thorndike (1934) advised that the first time a word is met is not ordinarily the proper time to learn it thoroughly; it might never again be met, or need such learning, or it might better be postponed to a time when the word has higher utility and less likelihood of being forgotten. "In increasing a pupil's reading vocabulary, we must select not only among words and among the various meanings of words, but also among the various degrees of accuracy and adequacy of knowledge" (p. 9).

3.4 *The Characteristics of Frequent Words*

The 10 most common words in a sampling of adult correspondence in American English around 1910 were *the, and, of, to, I, a, in, that, you, and for* (Ayres, 1913a, pp. 8–9). The 10 most common words in a sampling of American English printed in 1961 were the above, in virtually the same order, with three exceptions: *For* had dropped from the 10th to the 11th position, *I* had dropped to number 20 (to be replaced in the "top ten" by *he*—a fact to be reported to David Riesman), and *you* had disappeared from the "top twenty," probably because *you* is a word of correspondence more than of reading (Francis, 1967, p. 137). Stability, more than change, is a characteristic of commonly used words. That and their other properties will be discussed briefly. Each one of the indicated characteristics of high-frequency words has, of course, implications for education.

Grammatical words predominate. Linguists distinguish between grammatical words and lexical words. Certain representatives of both forms are in constant demand, not because of a given content requirement, but because of the very requirements of the structure of the language; these are the words of highest frequency across many studies (Bloomfield, 1933, p. 277). Grammatical words are also called "service words," or "function words," or "form(ative) words" because they consti-

tute a linguistic framework within which meanings are hung; being necessary to make up the structure of modern English prose, they are predominant among the very highest frequency words. The Kučera-Francis computational count found the most frequent 22 words, all function words, to constitute with their repetitions nearly one-third of the total of running words, that is, nearly one of every three words you will read is one of these 22 words. The first lexical word in frequency was in the 64th position; it was the word *new* (Francis, 1967, p. 134). (Social and cultural historians take note.)

It has frequently been observed that the service words, like abstract words in general, are harder to learn and harder to remember. Any teacher of reading in the primary grades can testify that *elephant* easily becomes a known "sight-word," while *with* does not. This empirical generalization is borne out by research in learning (Morgan & Bonham, 1944). Service words fail to interest students, Dolch (1951, p. 80) admitted in his textbook on the teaching of reading, but we teachers must teach them anyway for their essential position in the language. By gathering striking statistical evidence of these words' frequencies, vocabulary researchers did not invent the need to teach them—the language did that. But the need was underscored, with a probable increase in planned practice of such words.

"Culture-fair" words predominate. The conclusion drawn from the vocabulary research that established which are the most common words was that "differences in vocabularies due to sex, geographical regions, intelligence levels, and other factors of this type apparently do not affect significantly the base vocabulary of the 3000 to 4000 most common words. They may influence the marginal areas of the vocabulary, however" (Greene, 1941, p. 453; 1950, p. 389). Subsequent events have focused a great deal of attention and controversy upon the question of how "marginal" the sex linked, class-linked, and ethnic subcultural lexica are for education. The basal readers prepared since 1960 feature an ethnically integrated, less middle-class and suburban cast of characters. Reportedly based upon study of language usage in the inner cities, the earliest of this new genre, at least, demonstrate conformity to the orthodox reading view of a controlled vocabulary of words based on meaning and repetition and guided by word-frequency data; the Detroit City Schools series, in fact, presents fewer different words than the usual preprimers—a fact that increases the repetitiveness of words and the "opportunity for practice" (Barton & Wilder, 1964, p. 388n; Chall, 1967, p. 44).

Frequency counts were done locally in many parts of the United States and failed to show much geographical variability among the more commonly used words (Horn, 1939). Thorndike (1921) pointed to the fact that a list of generally useful words needs augmentation for pedagogical purposes:

The *Teacher's Word Book* does not, of course, rate correctly for any community, words which are very important locally (as, for New York City, *subway*, *elevated*, *Brooklyn*). By its very existence, however, it stimulates the educational authorities to extend and amend it in respect to words of special local importance. (p. 358)

The failure to find significant cultural variability reflected in the "essential vocabulary" must have been welcomed by the textbook industry. In order to recover the high development, production, and promotional costs of textbooks, publishers depend upon a national market, advertise in nationally read professional journals, and promote the news that large numbers of states and communities "have adopted our new series." The tendency toward sameness of content and the increasing trend toward a uniformity of their language has made textbooks a force for the standardization of American education, but by no means the only one (Clifford, 1975, pp. 34-40). The influence of New England's schools and educators on the rest of what became the United States operated in this direction almost from the beginning of our history. The emergence of school-profession groups in the 19th century set the stage for their expansions in the 20th. The National Education Association had 2,322 members in 1900 and 187,645 in 1935 (Evenden, 1938, p. 36), in the interim having appointed a number of very influential committees and commissions on goals and curriculum. They, like the *Yearbook* committees of the National Society for the Study of Education, were important in promoting a "conventional wisdom" to which vocabulary studies certainly contributed; word researchers were often represented on such bodies. The restriction, through language, of school content to a clearer, more culturally undifferentiated core was, after all, "appropriate" to a system of mass and universalizing instruction.

Cross-national equivalents predominate. The first frequency count of French words was published in the United States in 1924 by V. A. C. Henmon of the University of Wisconsin and an early student (PhD, Columbia University, 1905) of Thorndike's; only 400,000 running words were counted, so the results were accepted as only tentative. Nonetheless, and as Thorndike also expected, Ernest Horn (1926, p. 189) surmised that fuller counts would show high overlap between foreign-language frequency counts and *TTWB* and other English word lists. In an address before the National Conference in Research in English, in 1939, Horn confirmed the striking agreement on the relative frequency of English words and foreign-language words of equivalent meaning. "For example," he noted, "*time* is the commonest noun in the written vocabulary in English and *Zeit* is the commonest noun in German" (p. 132).

Buchanan's *A Graded Spanish Word Book* (1927) and Vander Beke's *French Word Book* (1929) were actual word counts in these languages. But Charles Ward's *Minimum French Vocabulary Test Book* (1926) was a transla-

tion into French of the most common words in English reading and writing. The results turned out to be in substantial agreement with the results of the actual counts. The achievement of comparable frequency findings tended to support the claims that frequency data possess validity and reliability. The international status of *TTWB* appears to have been confirmed by the fact that this was the almost invariable criterion against which the early foreign-language counts were measured. Many studies in England and the Commonwealth used *TTWB*, especially in learning research. An example of a less common overseas use of *TTWB* was Kelley's (1954) comparison of the vocabularies of primers used in Scotland; studies of this type were, of course, commonplace in the United States. The Thorndike *Word Book* is also being used today in research reported in the journals of non-English-speaking countries (*Social Science Citation Index*, 1974).

Many investigations were pursued out of curiosity about relationships between languages in word usage and out of the keen awareness of the vocabulary-learning problems posed by the foreign-language curriculum, in order to use science to establish the minimum essential vocabulary for introductory study. In 1934 vocabulary research was described as the most productive effort within foreign-language teaching (Coleman, 1934). It was reported that 60% of the foreign-language studies pursued in the period 1930 to 1937 were vocabulary investigations (Coleman & King, 1941, p. 528). Given the prevailing practices of a reading-translation approach, it was considered important to verify pupils' knowledge of the English equivalents of the foreign-language words being taught, and *TTWB* was used to ascertain that these were indeed common English words (Baker, 1928). Johnston (1939) found that 35% of the nearly 4,000 most common Spanish words were cognates of the words of corresponding frequency in *TTWB* and similar lists; around half of these cognates had the same frequency rating, in terms of groups of thousands, on the English and Spanish lists. As a final example, foreign-language specialists were caught up in the general movement to use frequency data to evaluate school reading materials, to grade books, and to secure simplified text materials. Using the Vander Beke frequency list as a criterion of difficulty, it was reported that, with 67% "easy words" (i.e., words in the first 2,000 in frequency), *Bataille de dames* was more worthy of teacher consideration than *Les trois mousquetaires*, with only 48% of its words "easy" (Hamilton, 1932).

Durable words predominate. A recent review of the Dolch list (1936) commented on its important historical role in helping to determine the vocabulary content of books for beginning readers and as a list of 220 "sight words" that many schools and teachers have made a point to teach. Along with other considerations, the age of the Dolch list raised questions about the need for its retirement (Johns, 1974). Similar queries were

raised before, such as Ruth Strickland's (1962) more general question of whether the vocabulary of modern school books has kept up with the language of their clients. Children reared in the 1950s and 1960s, she suggested, may use longer sentences than in the 1930s, perhaps because of greater verbal stimulation through the mass media, travel, and greater language permissiveness at home and in school (pp. 3-4).

Languages do change as new words are coined and others shift their meanings and gain or lose importance. Thorndike pointed out in 1934 that *gas* was once a scientific term known to few; in *ATWB-2* it was among the 2,500 most used words (p. 8). Even the most up-to-date list may be somewhat out of date; in the year that *TTWB-1* appeared, Thorndike (1921) wrote in the *Teachers College Record* of the "notable lag" that one must acknowledge:

It may well be that *automobile*, *auto*, and *Ford*, are now read oftener than *horse*, though the credit sum on our list is 108 for horse to 33, 13, and less than 10 for the other three respectively. The lag in school readers is perhaps even greater than that in our list. (p. 369)

The proponents of concentrating the language of the school upon the most important words, augmented by words of special value, would say, however, that high-frequency words have indeed "stood the test of time." Of a sample of words drawn from his list of 10,000, Ernest Horn (1939) remarked, "More of these words came into the language before 1099 than have come into the language since 1799" (p. 134). Thomas Horn (1969, p. 1284) reports a comparison of the words in Ernest Horn's *A Basic Writing Vocabulary* with the vocabulary of the correspondence of Benjamin Franklin (1706-1790): 97% of the words used by Franklin 10 or more times, and 95% of the words he used 5 to 9 times, appeared also in Horn's list. Given the high representation of the grammatical forms among the words of most frequent use, it should not be surprising that they are well represented among the oldest in the language.

Multiple-meaning words predominate. In the later 1920s and increasingly in the next decade, word-frequency researchers and educators had wondered whether word *meaning* was receiving sufficient attention, either in research or in pedagogy. In 1927 Dolch commented upon the irony of the inattention of most vocabulary researchers to the fact that high-frequency words have more than one meaning, when reading was coming to be known as primarily a "thought-getting process." The presumably simple words of primary-grade reading, including the service words, could pose a greater semantic challenge to children than their encountering a so-called technical word of but one meaning—as Luella Cole cautioned geography teachers in reminding them to attend to their pupils' total comprehension needs (Pressey, 1933). Early research by Fennell (1928) and Dearborn (1929) provided rich examples of confusions of meaning. Fennell illustrated the unwitting errors that pupils are led into

by their teachers. One child confused *just* as in "just gone" with the *just* in "He is a just man." Having had the first usage explained once by reference to the word *little* (i.e., "a little while ago"), the child refused to believe that the man in question was not "a little man." Lorge (1944) used the example of the child come home from Sunday School believing that the hymn he had learned had been about "the cross-eyed bear" (p. 549). It was pointed out that *in* has upwards of 20 meanings; some words have a hundred or more meanings (Seashore, 1933).

Readers' responses to multiple-choice questions given to test their knowledge of meanings consistently showed that performance in word recognition remained ahead of word-meaning knowledge all through the grades. The differences became greater the more one met the words of lower frequency (Gates, Bond, & Russell, 1938, pp. 166-167). In writing of the relationship of vocabulary development to better reading, Thorndike provided some illustrations of the variability one encounters when facing the semantic problem. Some words (like *seventy*, *oxygen*, *unicellular*) have relatively few and stable meanings; others (like *negative*, *cell*, *exponent*) have sharply differentiated meanings that emerge from context; the meaning of others (*good*, *bad*, *cause*, *natural*) requires both context and experience in divining among subtle variations in meaning (Thorndike, 1934, p. 7). This paper of Thorndike's may be read as the unofficial announcement of the new word work, on semantics, that he began formally early in 1934. The purposes and educational implications of the project, envisioned by Thorndike and Lorge, were stated by Lorge (1937) in this way:

The English semantic count is concerned with the discovery of the frequency with which each of the various meanings of multi-meaning words are used in a representative sample of English and American writing. Such a count will greatly supplement the straight frequency counts now available by enabling the writer and the editor to choose not only words which are relatively frequent but more specifically to choose the most frequent meanings of those words. . . . The semantic count will enable the teacher to do at least two things: To teach the most important words, and, to teach the most important concepts of these words. Since a sound educational psychology cautions against overloading the learner with materials that are not essential, concepts or meanings of words should be graded for importance, so that the teacher may teach them in order of importance or need. (p. 77)

The Lorge-Thorndike Semantic Count of English Words was included in *TTWB-3*. The data acquired for the semantic investigation showed a positive correlation between a word's frequency and the number of recognized (dictionary) meanings it possesses; a notable exception to this general relationship is the numerals. The positive relationship of number of meanings and the age of the word in the English language was also demonstrated here (Thorndike, 1948). Investigators, as they had done previously with frequency data, applied the semantic count to school

materials. Bachmann (1944) tried to determine the "meaning load" of two books written for children; some of the extremely common words of multiple meanings were *made* (25 meanings), *broken* (15), *fall* (14), *touch* (11), *bear* (10), and *right* (9). Thus the challenges to teachers lay with both the uncommon words' rarity and the common words' richness of meaning.

4. ON HOW AND WHY THINGS CHANGED IN SCHOOLS

Change in schools meant two sets of actions. One was the alteration of the reading materials favored in schools with the intention to improve their readability. The second was the concomitant heightening of awareness and acceptance of the opinion that learning words is an important and difficult task whose accomplishment cannot be left to chance. This awareness was manifested in many ways, including augmenting the treatment of words in the textbooks, using workbooks for word-building purposes, and testing students to verify a judgment about their word knowledge—the use of tests sometimes raising still higher the instructional emphasis given to words (Haggerty, 1918).

The purpose of this concluding section is to draw together, from what has gone before, generalizations about the forms through which Thorndike's primary contributions to change were made and to advance some propositions about the conditions or factors which appeared, together, to define the scope of that change. However important the Thorndike arithmetics, spellers, and dictionaries were to all who used them in schools, they will be left aside as proof of his influence, in favor of indicating the more indirect but pervasive stamp left upon standardized tests and schoolbooks.

The experiences of Thorndike and his associates in vocabulary research illuminate the existence of variables that facilitated and supported the direction of the indicated change. That the successes of the vocabularists were by no means complete, or uniform across the grade levels and curriculum areas, suggests that the factors of receptivity, resistance, or disinterest merit some discussion. While the historical *constellation* of conditions and forces was specific as to time and place and identity of the participants, there may be some general enlightenment of the processes of institutional change in this case study.

4.1 *The Media of Thorndike's Influence*

The standardized test. Truman Kelley once called social studies teachers' attention to the consequences of inattention to words in creating tests: "To employ the word 'indescribable' (having a frequency value . . . in the eighth thousand in the Thorndike *Word Book*) when testing for knowledge of 'church' (. . . in the fifth hundred in the Thorndike *Word*

Book) would pervert the test, it becoming a test of knowledge of 'indescribable' and not of 'church'" (Kelley & Krey, 1934, p. 137). As the making of tests and scales created a veritable deluge of new reading material after 1920, it was predictable that the selection and arrangement of test words would be subjected to something of the same sort of scrutiny that the words of textbooks were receiving. In fact, however, relatively few studies of test vocabulary appeared in the educational literature—even as the numbers of students meeting the tests grew by the thousands and then hundreds of thousands. Dale's bibliography (1949, p. 95) lists but 12 studies of the vocabulary contents of tests between 1923 and 1943. Applying readability measures to tests is more difficult than is using them with connected prose (Klare, 1963, p. 96). More recently, following the appearance of a readability formula designed expressly for the disconnected prose of most tests (Forbes & Cottle, 1953)⁵ and given the intense and highly politicized discussions of standardized tests, especially of intelligence testing, a new flurry of articles has appeared (Chall, 1958a). With vocabulary knowledge being, for example, an unrecognized factor in the vocational-interest tests being used for counseling junior and senior high school students, the validity of these instruments was seriously compromised when the test takers were students with poor reading skills (Roebor, 1948; Johnson & Bond, 1950; Cureton, 1951; Englehart & Beck, 1963).

Despite the generally low visibility of the test-vocabulary issue among educators, there is substantial evidence that test authors and test publishers have been considering and using the frequency word lists, and later the readability formulas, since the early years of the standardized testing movement (Otis, 1916). A survey of the users of the Dale-Chall readability formula found publishers of psychological tests among the users (Chall, 1956, 1958a, pp. 171-172). The recommendations of readability experts have been followed, at least in some cases (Forbes & Cottle, 1953). Research typically reports that, if the contents of the test are sometimes beyond the reading level of the average intended user, the test's *instructions* have at least been subjected to language control (Johnson & Bond, 1950; Forbes & Cottle, 1953). But the best evidence of the use of frequency data in general, and of *TTWB* in particular, is that found in the various test bibliographies, and especially in the highly regarded *Mental Measurements Yearbooks* (Buros, 1941, 1953, 1959, 1968).

Two lines of use demonstrating Thorndike's impact upon test vocabulary are evident in test reviews and test bibliographies. One is the use of *TTWB*, his dictionaries, or other "Thorndike words" in test construction;

⁵ To determine the readability of the language used in tests, the Forbes-Cottle formula employs the 1942 *Thorndike-Century Junior Dictionary* to assign weights to samples of words used in tests—if the words are above the first 4,000 in frequency (Forbes & Cottle, 1953, p. 186).

such use appears as early as 1913 (Wang, 1939). The second application of Thorndike words is as an instrument used by reviewers to comment upon the appropriateness of the language used in tests—including spelling tests (Buros, 1941, pp. 139–140) and clerical tests (Buros, 1953, p. 717). The same considerations about word selection and order of test items, with reference to *TTWB*, appeared in both lines of application.

Intelligence and scholastic aptitude tests, with their vocabulary subtests, are predictably places where Thorndike's work would, and did, have obvious use. As early as 1926, Schwesinger stated, "The appearance of Thorndike's *Teacher's Word Book* in 1921, marked the initiation of a change in the method of selecting and constructing word knowledge tests" (p. 4). Aside from his own participation in devising the National Intelligence Test (1920) and CAVD (1925), one of the more recognized tests that made use of Thorndike's frequency-count data was the Guilford–Zimmerman Aptitude Survey (Guilford & Zimmerman, 1948). While the accepted theory of selecting alternatives (wrong answers, distractors) was the one clearly discussed by Thorndike in 1916 (and in 1925 by Thorndike and Symonds), Guilford and Zimmerman made the "difficulty" of the distractors equal to that of the stem word ("test word," "stimulus word"), on the theory that this reduced the processes-of-elimination of knowing the easier distractors and was, therefore, a better test of a person's knowledge of the word being tested; a correlation of .87 was found between the arrangement of words for difficulty (frequency), based primarily on *TTWB*-3, and empirical trials of the test.

Tests of reading, spelling, and vocabulary employed the frequency placement of Thorndike's or other lists for several purposes: selecting the words that comprised the test content, choosing words to act as distractors, and arranging test items in order of ascending difficulty (Schwesinger, 1926; Gansl, 1939; Davis, 1944; Kirkpatrick & Cureton, 1949; Fries, 1950, p. 23; Miles, 1973). Validity checks against *TTWB* were reported as a "very useful" practice in test construction (Rinsland, 1937, pp. 291–292). The O'Connor Junior English Vocabulary Test (O'Connor & Filley, 1933) and O'Rourke's Survey Test of Vocabulary (1931) were early examples (Gansl, 1939). Coming later were, among others, the Nelson–Denny Reading Test (Buros, 1941, 1968; Horrocks & Schoonover, 1968), the New Standard Vocabulary Test (Buros, 1959), and the Seashore–Eckerson General Recognition Vocabulary Test (Seashore & Eckerson, 1940). The use of the *TTWB* and the Semantic Word Count in the construction of the very important Cooperative English Tests, published by the Cooperative Test Service of the American Council on Education, is well known among psychometricians; comparisons of the correlations of *TTWB* frequency ranks with the empirical difficulty of test items gotten in this and other tests stimulated some thinking about "range" of vocabulary knowledge as compared to "preci-

sion" of word knowledge as a goal of vocabulary testing (Buros, 1941; Davis, 1944; Kirkpatrick & Cureton, 1949).

Teachers of arithmetic have demonstrated concern about the readability of arithmetic materials through this century, and arithmetic books showed attention to control of vocabulary difficulty ahead of some of the other curriculum areas. It may be that this history of word-awareness was what led the makers of the arithmetic test of the Sequential Tests of Educational Progress (STEP) to attend to vocabulary to an unusual degree. In the words of reviewers R. L. Thorndike and E. Hagen (1961), "In a commendable effort to incorporate the arithmetical tasks in real and meaningful problems, the authors have introduced a reading level such that the resulting score has a higher correlation with a verbal than a quantitative score on the parallel aptitude series (SCAT)" (p. 308). This contrasts markedly with the observation that E. L. Thorndike made in *The Elementary English Review* in 1940 on the fact that the instructions printed on certain tests of vocabulary contain such rare words that they are extremely hard to understand.

Test construction also went on outside the major centers of test production; a brief reference to Thorndike's influence upon such tests should conclude this discussion. The several books published for the needs of schools and teachers in the construction and evaluation of better tests almost invariably speak of appropriate language and careful attention to vocabulary in the content and test directions. Such remarks, even when accompanied by examples, are so general that they can be counted only as making evident the extent of awareness of readability issues. Specific reference to use of *TTWB* is not lacking, but it is widely scattered and difficult to acquire. Nonetheless, the public secondary schools of East Chicago, Indiana, published an account of their own test of Social Science Vocabulary for secondary-school use. Locally constructed and in use for several years at the time of this report, the tests used words selected from the Kelley and Krey list, "Terms in the Social Sciences" (Kelley & Krey, 1934); the terms were arranged in the test in order of the words' frequency of use, as determined by the *Thorndike-Century Junior Dictionary*, and the definitions and distractors were words selected on the standard principle of being "easier" (i.e., more common) than the stimulus word (Bolton, 1943). If such procedures were used in locally constructed tests in smaller communities, they certainly were in the largest urban districts that maintained test and research bureaus staffed by the graduates of Chicago, Iowa, Teachers College, and other important centers of research and graduate studies.

Public schools not only gave tests by the millions of copies to their own students but were sometimes charged with administering tests to other populations to meet requirements of the community or state. The situation in New York State in the early 1920s illustrates such a function. New

York was one of the 19 states that had passed a literacy-test requirement for new voters. The 1922 revised election law provided for two means of showing proof of literacy for voter-registration purposes: (a) a signed testimonial or certificate of literacy from a public school administrator, or (b) "intelligible" oral reading and legible writing of 10 words about what was read, the reading matter being randomly selected 100-word extracts of the state constitution. A committee of educators and specialists in educational measurement was formed to formulate a test to assist in meeting the first option; in the process it submitted the state constitution to examination by *TTWB* and found that only 1,120 of the 2,098 different words of the constitution were in the first 4,000 words and that 658 were not in *TTWB* at all. The result—after revision of the election law, a court test, and construction of a standardized test under the authority of the Regents—made the State Regents Literacy Test the only test for new voters, with the schools being given sole power to administer the test. In the first year of its administration, 1923, over 28,000 new voters qualified; of these, 27.5% did so by having day-school credentials, 2.7% by having attended night school, and 70% by passing the examination. The percentage of test-passers was 79%, while 21% failed. Of the struggle, more politically difficult than scientifically demanding, a participant exulted:

Thus the highest courts of the Empire State upheld the fundamental principle laid down by the committee, namely, that a test of reading ability should measure the candidate's ability to read with comprehension of the meaning of what he read. Thus the work of the makers of standard silent-reading tests has reached over into the field of civic life far enough to induce a leading state to change its literacy requirement from ability to "read intelligibly" to ability to "read intelligently." Since our states presume to require educational qualifications for new voters, New York has demonstrated how such a test may be scientifically constructed and intelligently administered. (Morrison, 1925, p. 155)

The omnipresent textbook. Three *Yearbooks* of the National Society for the Study of Education appeared between 1920 and 1931 inquiring into the suitability of school materials for meeting the challenges and opportunities confronting American public education. Comparisons of topical coverage of different subjects had been made in most of the curriculum fields. There were even a few small studies of the degree of overlap in vocabulary from one textbook to its competitor. There were occasional references to language difficulties and "hard words" in the journal reviews, and assurances of "suitable language" or "simplicity of style" in advertisements. With *The Teacher's Word Book* came a new criterion, but one built out of the old and, in this culture, valued principle of "social utility": How does textbook language compare with the general language using word frequency as the measure? It appears inevitable that Thorndike's *Word Book* would be widely used, given the quest for precision that marked many professionals in that era, both within and without

the schools. In *The Materials of Reading, Their Selection and Organization*, Willis Uhl (1924) wrote of *TTWB*:

The value of this word book lies chiefly in its use as an objective measure of the appropriateness of the vocabularies of school books. In so far as the sources are representative and adequate, this list of words enables school men to judge one of the fundamentals of content—the vocabulary. Additions will be made to the list, and the order of occurrences of the words will probably be changed. But until additional investigations are made, the choice of textbooks should be influenced by standards based upon this study. (pp. 25–26)

The “additions” made were largely those of Luella Cole and others working at more definitive lists of technical words. The “additional investigations” conducted were made by Thorndike and his associates and appeared as *ATWB*-2 and *TTWB*-3.

The implications of the ability to affect, even in apparently small ways, the contents of America's school books may be made clearer by quoting Wilbur Schramm (1955):

It may well be that the most influential group of men in this country in the making of school texts are the educational editors of our large publishing houses. Their only real competitors are the ideological leaders like Dewey and Thorndike whose concepts of what a textbook should be and how it should be made have been reflected in the efforts of hundreds of editors and authors. (p. 158)

American textbooks have been occasionally praised, as when it was written that “the American publishers are among the great educators of the world” (Hall-Quest, 1918, p. 10). Ernest Horn (1937) argued that criticisms of textbooks lacked perspective: “In comparison with typical courses of study, the average textbook is superior in scholarship, pedagogical arrangement, and the methods of teaching that are recommended” (p. 217). American textbooks have also been damned. Wayne Otto (1971) has expressed a view that has been heard often in this century, referring to “the despotic stranglehold of textbook publishers” (p. 448). Above all, American textbooks have been bought (Cronbach, 1955, p. 19; Mayer, 1962).

Many American educators do not realize that textbook use is more prevalent in this than in other systems—even though the contrast (probably overstated) has been noted at least since the 19th century. The contrast with their own more oral methods caused some Europeans to label the textbook method as “the American method” (Horn, 1937, p. 207). In Germany, most texts, where they were available, were largely outlines which the teacher filled in (Hall-Quest, 1918, p. 40; Learned, 1927, pp. 13–14).

The reasons given for American dependence upon the textbook centered on two views, one sanguine and cultural, the other pessimistic and social-historical. Teaching by the book is, as Horn (1937) put it, “chiefly

accounted for by the conviction that the pupil should develop the ability to study independently" (p. 206). One of the reasons that the Hamtramck, Michigan, schools gave for developing a local wordbook—checked against *TTWB-1*, Gates's, Horn's, and Packer's lists—was that "children would have more independence of the teacher; they could look-up a needed but unknown word, as well as add their own words to the list as the school year progressed" (Watters, 1936). Not inconsistent with this explanation is the historical fact that American teachers had been comparatively untrained and impermanent in their posts for most of our national existence. The German oral teaching method presupposes a degree of teacher professionalism that most American schools would have found amazing; as late as 1918, more than half of America's elementary-school teachers had two years or less of academic and professional training beyond the high school (Evenden, 1938, pp. 37–38). In this context, textbooks often functioned as undeniable crutches (Hall-Quest, 1918, pp. 2–3), and some educational leaders have yearned for, somehow, a "teacher-proof" curriculum. In a monograph on *The Vocabulary of Arithmetic* (Buswell & John, 1931), the authors wrote of the possible ways of presenting new technical terms in arithmetic:

First, they may be presented and explained by the teacher, who can see that experiences are provided with which the new term may be associated. Second, technical terms may be introduced and explained in a textbook. If teachers were uniformly well-trained, the major burden of introducing new terms might be placed on them. In view of the limited professional preparation of many elementary-school teachers, however, the chance of introducing new terms successfully is increased when the textbook carries part of the obligation. (p. 91)

In a similar vein was an argument advanced for a light vocabulary load in primary-grade reading materials, that, thereby, less teacher expertness is required for success in teaching children to read. If books are easier, the less experienced teacher succeeds; if books carry a heavy burden of words to master, even the experienced teacher may fail (Hildreth, 1948b). Admittedly, this was written during an aggravated teacher-shortage when "emergency credentials" were issued to the manifestly unprepared. Nonetheless, doubt about teacher competence has been openly and freely acknowledged for a very long time in America.

Perusal of the journal literature in education of this century, especially of the period 1920 to 1945, makes clear that the pressure for more readable text material was not the publishers' only concern. The question of free textbooks continued to agitate public opinion, as it had done since before Massachusetts passed (in 1884) the first of the 15 statewide, mandatory free-textbook laws that existed in 1918. A mass of permissive legislation, regulation, and qualification had appeared. The system of "adoption" of uniform textbooks might be applied statewide in some areas

and in some subjects, while local adoption prevailed elsewhere. The adoption system generated its own sizable literature on rating systems for textbook selection (Franzen & Knight, 1922; Harap, 1934), as well as being a subject of frequent debate and even conflict. It was also a lucrative system, provided one's text or series got adopted in one of the big state or city systems; that often started a landslide of imitation. Gertrude Whipple surveyed state and local procedures of adoption in 1936, finding a very small percentage of plans requiring statistical measures; if the textbooks' authors or publishers had already resorted to readability strictures, however, this finding becomes much less important. Chall (1956) reported that school systems were not heavy users of readability formulas for selecting or grading textbooks; teachers reported making their own judgments of vocabulary difficulty or accepting the publisher's grading.

Censorship battles over textbooks came up often, sometimes as part of a general paranoia—during the “Red Scare” of the early 1920s and the antisubversive purges of the 1930s—and other times as the result of some local incident. The most celebrated case concerned the successful social studies textbooks of Harold Rugg; dozens of articles were published in defense of one or another principle. In the 1930s, however, more schools faced poverty than faced persecution. The textbooks were reported as falling apart, and more school districts calculated rebinding costs than calculated readability—it appears safe to say. A vice-president of one of the largest publishing firms pointed out that the Great Depression was the first depression in American history to hurt textbook sales: Sales by 1935 were off by 35%. In previous depressions, rising enrollments had kept sales up; in previous depressions most books were bought by parents, not provided by the schools (Nelson, 1935).

In spite of other severe pressures, the challenge of vocabulary was not a secondary one to the publishers. A threat to their sales was both implied and articulated, as Edgar Dale (1931b) showed:

We are inevitably led to the conclusion that the research worker in the field of vocabulary must lay rather definite plans for the utilization of his results if he hopes to have them affect school practice. Perhaps the most effective way of doing this would be to refer to the textbooks analyzed by name, instead of by code number, as is so commonly done. Nothing will affect the vocabulary of textbooks so much as widespread publicity to the data secured by our analyses. (p. 122)

Theirs is a highly competitive business and the fact remains that publishers' claims to having produced a textbook or series of textbooks “constructed according to the latest scientific findings” *did* sell more books. One of the first prominent educators to note this, in one of the important journals, was Carleton Washburne—nationally known superintendent of the Winnetka (Illinois) schools, an early and enthusiastic user of *TTWB* to grade books and confirm the validity of spelling lists, and a word-counter

in his own right (Washburne, 1971, pp. 467-468). The "phenomenal success" of a speller based upon scientific investigation was, he wrote, evidence of "how eager the educational world is for curriculum material based on scientific study rather than on personal theory" (Washburne, 1923, p. 751).

Even though vocabulary control was one of the factors increasing the costs in a highly competitive industry, the publishers' "market research" mandated meeting the competition (Schramm, 1955; Gates, 1961; Chall, 1967)—first and most notably in the primary reading textbooks but eventually, although to greatly differing degrees, in all curriculum fields (Gray, 1929, p. 449; Beust, 1951; Yoakam, 1954; Chall, 1956, 1958a; Anderson, 1958, p. 183; Mayer, 1962; Smith, 1965, p. 280; Bormuth, 1966, pp. 81-82). Textbook authors were increasingly drawn from universities' departments of education and from educational research (Richey, 1931). Prominent vocabularists or their students authored many of the major publishers' textbooks and series, especially in reading, spelling, and elementary-school mathematics; one thinks immediately of such men and women as Horn, Gray, Buckingham, Dolch, Yoakam, Betts, Powers, N. B. Smith, Gates, and Hildreth.

The Committee on the Textbook of the National Society for the Study of Education reported in 1931 that it found "in the prefaces and advertising materials for textbooks in various elementary-school subjects the claim that scientific techniques had been systematically applied in creating the textbooks" (Schorling & Edmonson, 1931, p. 27). Exaggerated or meaningless claims were undoubtedly made, out of promotional pressures or misunderstanding and enthusiasm. Mallinson (1952) advised science teachers to be cautious in accepting the claims of publishers and authors about vocabulary load, mentioning a recent text's claim to contain "only three hundred fifty words of new technical vocabulary"; such a number, he pointed out, may be too large if the "normal" increase in all vocabulary for that grade level is 1,200 words, too small if it cannot explain adequately the intended scientific meanings, and include many nonessential words (p. 273). Still, Yoakam reported that, by 1945, in school reading books at least, the publishers' placements accorded rather well with readability levels using his formula (which applied *TTWB*), whereas a decade earlier publishers were overestimating the readability of their books. A major difficulty in assessing the precise nature of the publishers' applications of vocabulary and other readability research to school textbooks was the secrecy of techniques used in appraising and rewriting manuscripts. The NSSE's textbook committee pointed out, in its investigation of publishers' procedures in controlling the vocabulary of arithmetic textbooks, that "the first method is to check against the *Teacher's Word Book* by Thorndike or some word list that is the confidential property of a publisher" (Schorling & Edmonson, 1931, p. 39). As readability formulas

evolved and multiplied, the commercial value of a successful in-house variation increased. Chall (1958a, p. 113) came to call readability measurement almost a "secret tool." This suggested to Klare (1963, p. 93) the reason why such a widespread application of readability research has not generated a larger literature.

Textbooks promoted as having a scientifically determined content obviously exceeded research findings (Ayer, 1945; Chall, 1967, p. 194n). *TTWB*, for example, was not intended as a spelling list; Thorndike stated clearly in the *Word Book* that his words were listed by lexical entries (*long*, *longer*, *longest*, all included with *long*) and not by inflectional forms, where many spelling problems arise. Nonetheless, spellers were constructed from *TTWB* and such "misuse" was noticed and commented upon in the educational journals by other researchers (McKee, 1937; Seegers, 1939). Textbook advertisements projected a confidence in the definitiveness and predictability of readability research that certainly strained warrantable limits: "A vocabulary well within the comprehension range of students in the secondary schools" (geography text, 1938); "vocabulary and sentence structure insure ease of reading" (arithmetic text, 1938); "written in the vocabulary of the grade *lower* than the grade in which it is to be used, insuring easy, successful reading for the average child" (readers, 1974); "word introduction and maintenance grounded in semantically sound situations with complete vocabulary control" (readers, 1948); "have been vocabularized according to Fifth-Grade Word Lists" (readers, 1948); "graded according to the child's reading ability" (science texts, 1954); "popular classics in the First Thousand Words for Children's Reading" (1962). The pre-1920 resort to general claims about "ease of reading" and "appropriate vocabulary" was largely supplanted by more and more specific statements about the grade placement of the vocabulary used to write textbooks. The omnipresence of graded textbooks prompted one set of investigators to ask, What do publishers mean by "grade level"?—and to check the publishers' claims for 200 textbooks against the results obtained by using readability formulas; in half the cases they disputed the publishers' grade-level placement (Mills & Richardson, 1963).

An indisputable fact about textbooks in the 20th century is the growth in the amount of reading matter they set before the student. President Eliot of Harvard University had calculated in 1898 that the entire 6-year elementary-school curriculum could be read in 46 hours, a commentary on the fact that the brief textbooks then in use were modeled after the catechism and intended to be memorized, not read; in contrast, Charles Judd noted in 1936, "Pupils in present-day elementary schools read ten times as much as Eliot found pupils reading in 1898" (p. 575). Thorndike had made the same point more fully in 1934, illustrating the deficiencies of traditional textbooks in meeting modern reading objectives and in satisfying the requirements of learning theory:

Apart from other inadequacies and difficulties, there was not enough reading. The series for Grades 4, 5, 6, 7, and 8 would contain at most about 500,000 words. About nineteen-twentieths of these words would be within the commonest 2,500. The 25,000 or so remaining would permit practice four times a year on 1,000 words, or twice a year on 2,000 words, or once a year on 4,000 words. If 8,000 words beyond the 2,500 commonest are included, there will be on the average one occurrence of each such word every two years! A year of the old *Youth's Companion* had many times as much reading as a whole series of Readers. A girl reading *Little Women* read nearly as much as she did in reading her Readers for Grades 4, 5, and 6 altogether. A child should read at least 5,000,000 words in Grades 4 to 9. (p. 5)

Given the need for adequate repetition of vocabulary and the greater specific information available about technical vocabulary, the textbooks in the content subjects also increased the amount of reading they required. In the social studies, the change was described as featuring "much more reading matter" (Murra, Wesley, & Zink, 1941, p. 1146). Even the harshest critics of modern high-school English anthologies commended "their quantitative sufficiency of reading matter" and generally adequate treatment of vocabulary study (Lynch & Evans, 1963, p. 409).

Along with the sheer increase in textbook bulk, additional exposure to words in print came with the appearance of supplementary books—a move which Thorndike endorsed as an alleviation of "the inevitable inadequacy of any series of Readers" (Thorndike, 1934, p. 5; Uhl, 1937, p. 211; Clement & Dolch, 1941, p. 1305; Smith, 1965). In reading, pre-primers began to appear after 1925, some preparatory to a given series' primers and others independent of any series. Within a quarter century, one found multiple texts everywhere in evidence—with associated readers, enrichment readers, simplified readers for slower students and additional readers for the faster, "adventure books," "sports series," "aviation readers," etc. In this flood of material it was possible to anticipate a recurrence of the considerable fluctuation in vocabulary practices that had characterized textbooks in the earlier 20th century; that this did not happen, that "great and unwarranted disparity among the vocabularies of books for children" was prevented, was credited to the influence of the Thorndike list (McKee, 1937).

Independent series like the "Walt Disney Story Books," published by D. C. Heath and Company in the late 1930s, reinforced the vocabulary of the basal readers by editing their content in conformity with standard word lists. The Macmillan Company published the "Good Companion Series" in the 1930s and the "Core-Vocabulary Readers" in the 1940s as supplementary to any basal reading series, on the grounds that they were constructed of the words most frequently appearing in the several most commonly used basal reading series—a claim also made for the Steck Company's "Woodland Frolics" supplementary readers. The appearance of graded series of basal textbooks in the content subjects of science, social studies,

and arithmetic allowed for easier vocabulary coordination across the curriculum. Not surprisingly, a study of the readability of various elementary science textbook series generally found that those published by the firms most successful in the basal reading field were the most readable of the science books (Newport, 1965). Graded series also successfully met the objection students directed against earlier ungraded textbooks, that content encountered in one grade would often be met again in a later grade (Uhl, 1921, p. 40).

Textbook expansion provided for the more systematic introduction and repetition of new and uncommon vocabulary. So did the workbooks that appeared and proliferated after 1926—a movement described as “one of the most striking modern trends” (Horn, 1937, p. 220) and as “almost phenomenal” (Murra et al., 1941, p. 1149). Workbook sales approached \$30 million by 1956 (Buckingham, 1960, p. 1520). McAulay's comparative study (1948) of geography textbooks available in 1938 and 1948 noted that workbooks based upon textbook content had become commonplace. One of their functions was to provide exercises to stimulate “vocabulary development,” a particularly common phrase after 1950. The value of workbooks as “a boon to inexperienced, incompetent, or overworked teachers” (Horn, 1937, p. 222) was probably less questioned than was their real contribution to realizing the oftentimes grandiose educational objectives so widely articulated during much of this century. The teacher's manuals, also grown in size and specificity, further identified vocabulary loads, probable sources of word difficulties, and means by which the teacher could enhance comprehension. Considering how much teachers had their opinions formed by the manuals (Chall, 1967, p. 188), their attention to vocabulary was crucial.

In 1935, Guy Whipple, an editor of a major educational publishing house, offered some autobiographical insight into the demand for more readable teaching materials:

About once in a dog's age (it has happened once in my editorial career) complaint is made that a textbook is too easy for the grade for which it is designed. Very common are complaints that a textbook is too difficult. Whether justifiable or not, the pressure for simplification seems steadily to increase. (p. 579)

Whipple expressed concern that readability was too often equated simply with verification of vocabulary against a word list. Less than two decades later, the existence of vocabulary control, as governed by *TTWB* or some similar list, was the staple of publishers' concessions to readability (Betts, 1949a, p. 449). At least insofar as writing textbooks for the primary grades was concerned, the editors of two other firms appeared to take such a practice completely for granted: “Vocabulary has to be checked and rechecked against proper word lists; new or special subject matter words must be introduced gradually; repetition of these same new words

must be looked out for, sentence length and difficulty carefully graded" (Livengood & Loveland, 1952). In a textbook on the teaching of reading published during the same period, a well-known author of a basal-reading series (Dolch, 1951) contrasted past and present practice and indicated the existence of seemingly rigid rules of procedures:

Time was when the author of readers for children just wrote what he thought appropriate in the best way he knew how. As a result, a typical case was a reading series in wide use in 1925 which had 2,500 different words in the primary books and 1,100 of these were used a single time only. Nowadays, the author writes what he thinks children should read, and then a battery of editors goes to work on the manuscript. (p. 170)

Typical rules applied by such editors, Dolch indicated, were that first readers introduce no more than three words per page and third readers no more than five, that every new word be repeated on the same or immediately following page, and that no word be repeated less than five times in a book. Adherence to the principles of vocabulary *selection* and *distribution* had been widely articulated and practiced in the reading field before 1935 (Charters, 1924, p. 135; Fennell, 1928; Wiley, 1928; Foran, 1934; Hockett & Neeley, 1936; Reed, 1938). The "Newson Readers" advertised, early in the 1930s, their adherence to the small-vocabulary, high-repetition principle: 50 different words in 1,242 running words with an average repetition of 24.84, this in their preprimers. The "Unit-Activity Reading Series" (1935), edited by Nila Banton Smith, advertised itself as having the "smallest total vocabulary—largest number of running words—greatest average repetition."

Dolch (1951) contrasted such "good practices" with the situation still found in some of the textbooks in the content subjects prepared for the middle and upper grades. Witness the science reader which presented the names of 22 small woods animals on one page—"an impossible learning task for the child" (p. 291). Equally ridiculous it was "for makers of textbooks to think that they have solved the 'strange word' situation by adding a glossary of terms to the back of the book, or by giving advice to 'use the dictionary'" (p. 292). The last word, however, on the progress made in extending considerations of readability from textbooks used in teaching reading to those of the entire curriculum of elementary and secondary schools goes to B. R. Buckingham (1960)—an early student of Thorndike's and then education editor and partner in one of the most successful publishing houses:

The evidence concerning the effect of readability studies upon textbooks is spotty and in general inconclusive. The period during which the readability studies were made runs from the middle twenties to the middle fifties. At the beginning of this period all were agreed that textbooks were too hard for the student who had to use them. Now, at the end of this period, it is safe to say they are still too hard; but there is good evidence for saying that they are not so hard as they used to be. (p. 1520)

Given the great many factors that go into determining what happens in public education, vocabulary researchers have some reason to feel consequential—even in teaching fields that were the most resistant to innovation, as determined by quantitative studies in education.

4.2 What Makes Research Influence Practice?

The pedagogical imperative of vocabulary teaching was frequently directed at teachers and textbook authors of the content subjects in the middle third of the 20th century. "Vocabulary studies have become so common that it is ordinarily assumed that any new materials of instruction will be carefully controlled in this respect," wrote the University of Chicago's Guy Buswell in 1933. "Familiarity with these vocabulary investigations is expected of any competent supervisor [and] the application of studies of this type to the construction and evaluation of teaching materials is too obvious to require any comment" (p. 343). Surveying vocabulary ordinarily used in 13 school subjects, much of it vocabulary outside the words of *ATWB-2*, Luella Cole (1940) also employed the imperative: "There is no subject so easy that all of its essential words fall within the 20,000 most common words in the language. Every teacher must, therefore, consider the teaching of vocabulary one of the fundamental requirements in the presentation of subject matter of all sorts" (p. 24). Thorndike (1936-1937) pointed out that knowledge of words above the first six or seven thousand most common would increase even more slowly without the content subjects than it does (with them). Indeed, he reminded teachers, "much of their [students'] learning of geography, history, science, and literature is learning the meaning of words" (p. 417).

Judging by students' word knowledge, however, the effect of such exhortations was mixed. In 1952 a Virginia teacher reported, as her master's thesis, the results of a vocabulary test administered to 684 fifth-grade students, finding that only in the arithmetic and health fields could students display knowledge of more than half the common technical vocabulary of the content subjects (Johnson, 1952). Whereas textbook terminology in arithmetic was correctly defined in 68% of the cases, only one science term in three was known; between these extremes, ranked from high to low, were the word knowledges of health, history, geography, and literature. In arithmetic no word was unknown to 90% of the pupils, but 90% failed to select the correct definition in a multiple-choice test of the following words in their respective fields: *arch* (health), *common* (history), *frigid* (geography), *reign* (literature), and *substance* (science). Evidence of a similar variability across fields is available in the testimony of subject-matter specialists, in studies of textbook vocabulary, and in content analyses of the pedagogical literature of the several curriculum areas. Explanation of the differential effect of vocabulary research will be attempted below—in an effort to cast light on the conditions of a teaching

field and the qualities of a research tradition that enhance or limit responsiveness to educational research.

Social sanction. As social agencies, public schools in the United States must be answerable, at least in part, both to basic cultural imperatives and to less permanent social forces. By the 20th century it was widely agreed that public education was a necessary social investment, that popular aspirations and national social and economic well-being demanded that it also be universal. Educational practices that were aimed at educating "all of the children of all of the people" had wide support, inside and outside the schools; those that smacked of elitism or exclusivism had to operate more obscurely to survive.

The schools by 1950 were receiving and, more importantly, retaining larger proportions of the less "word-facile" of the population. Thorndike (1936-1937) wrote:

The books that were written in the past were more or less suitable for the top third or quarter in intellect and ability to read. But to expect the total population at age 14 or even 16 to read books with a vocabulary of 40,000 words exclusive of names of persons and places, and including such words as calamanco, calash, calcareous, canaille, canonize, and canonry, is folly. (p. 419)

The economics of mass textbook publishing and school district purchasing promoted books aimed at the average reader, not the superior one—the result being a simpler textbook than formerly was the standard (Durrell & Sullivan, 1938).

Another prominent reading and vocabulary expert made even more explicit the relationship of readability to democracy in education: Those who espoused school books written to meet the canons of scholarship in the several disciplines, letting the pupils take their chances, Yoakam (1954) characterized as having "an aristocratic rather than a democratic theory of education." More democratic would be the preparation of books under the credo of the greatest good for the greatest number; "personally we believe in the principle of more adequate material for the average and less able child. The bright will take care of themselves," Yoakam believed (p. 430).

Lest this latter also be considered undemocratic, researchers could point to the fact that complaints by teachers and students of material being too difficult were well-nigh universal (Uhl, 1921). There was additionally the research demonstrating that better readers also profit by reading material from which unnecessary obstacles had been removed; retention, reading rate, and attitude were all found related to readability factors in one study (Klare, Mabry, & Gustafson, 1955).

Teachers added their perceptions to those of the vocabulary researchers, on behalf of books and courses of study that better fit the range of their students' abilities and backgrounds. "Teachers desire material which increases interest instead of material which needs so much explanation

that interest is lost by the time that meaning is mastered or the words correctly pronounced," Uhl reported after a large survey of teacher and student reaction to reading materials (1921, p. 57). Teacher testimony was secured in an early study of the suitability of the vocabulary of arithmetic textbooks. There, too, was evidence of clear teacher concern for excessive word difficulty and books otherwise inappropriate to the grade for which they were ostensibly designed (Chase, 1917). Luella Cole (1934) approvingly reported to prospective teachers the initiative shown by a teacher who sought to verify her own suspicions of the unsuitability of a textbook she was assigned:

One teacher a few years ago was impressed with the fact that the reader used in her third-grade room was too hard. She therefore made a list of all the words used in that and in the fourth-grade reader of the same series. She next compared the two lists with the Thorndike Word Book and was able to prove that the vocabulary used in the third-grade reader was actually more advanced than that used in the other. This teacher started with an impression; instead of simply talking about it, however, she proved its correctness in terms that others could understand. (p. 300)

In talking about simplifying school materials so that *all* children would have "a fair chance" in school, vocabulary reformers were speaking in terms that the larger society could understand and would appreciate—more so than were their critics who spoke of "lowering standards" or "debasing the language."

A "logical fit." It is hypothesized here that curriculum areas which are obvious "word fields" would simply be more responsive to vocabulary research than not. Spelling in the elementary school is obviously a "word field"; so are the commercial subjects of typing and shorthand in the secondary schools. However much their teachers might envision larger educational goals, mastery of the manipulation of the symbols which stand for concepts has to be their immediate concern.

Reading, at least in the initial stages of instruction, is also a word field; regardless of the stress placed upon meaningfulness and thought processes, the acquisition of a sight vocabulary or the articulation of pronunciation elements into recognizable sound packages commands the attention of teacher and pupil.

As suggested earlier in this chapter, the reading field illustrates a curriculum area predisposed to welcome vocabulary research and its implications. The principle of vocabulary control had been exercised in the post-Reformation versions of the Bible prepared for teaching children to read—although the Bible itself is "easy reading" in word-frequency terms (Dolch, 1938; Chall, 1958b). A process of simplification of the language of school readers was evident in the 19th century, perhaps an unconscious consequence of a significant cultural change in language behavior; Sherman (1893), for example, noted a marked decline in sentence length in

English, and an increase in the proportion of simple sentences employed, over several centuries. Preceding the word counts of modern vocabulary research, however, the reductions in vocabulary load, provisions for word repetition, and efforts at grading material for difficulty progressed in an unsystematic, empirical, trial-and-error fashion (Yoakam, 1945; Fries, 1950; Hunnicutt & Iverson, 1958; Smith, 1965).

Huey's *The Psychology and Pedagogy of Reading* (1908) clearly anticipated the emphasis that Gray and others placed upon meaning and reasoning in reading, upon the "subtle but essential aspects of reading [which] are not to be strangled by the pressure of drill on words" (MacLachy & Wardwell, 1948, p. 199). Primer vocabulary was reduced by about one-third during the 1920s and 1930s, from an average of 314 different words in primers published before 1926 to 207 for books published in 1936; basal series for the first three grades reduced their total vocabulary from 1,600-1,800 words in the mid-1930s to as few as 1,100-1,200 a decade later; over a quarter century, the vocabulary burden of basal readers was halved and provisions for word repetition grew markedly (Hockett, 1938; Yoakam, 1945; Sisters Dorothy & Cecile, 1946; Kearney, 1950; Chall, 1958a; Smith, 1965). Thus a new orthodoxy became firmly established in reading instruction. In the words of a committee on reading of the National Society for the Study of Education, "If children are to make joyous, consistent growth in reading, basic reading materials should be simple . . . and new words should be so paced that most of the words in a given selection are familiar" (Berry, 1949, p. 179).

Again, at least in its beginning stages, foreign-language teaching is another word field. Teachers of the classical or modern languages might, and did, argue about the relative emphasis that should be given to the cultural outcomes of language study, but the preoccupation with teaching beginners a basic vocabulary was very evident in the literature (Coleman, 1929, 1938; Hewitt, 1936; Cole, 1937). It is no coincidence, as Hildreth (1943) thought it was, that reading and foreign-language instruction entertained the same trend in their textbooks: low rates of vocabulary addition and frequent repetition of old and new words.

Thus, in each of the foregoing subjects, teachers were in essential sympathy with the efforts of vocabulary researchers to identify an essential word list of limited scope and to clarify the processes and parameters of vocabulary development. These activities "made sense"; their logic was apparent. Forced to be concerned with word mastery, the teachers of these word fields probably also had greater appreciation of the principle of repetition, drill, and practice—which was so central in the scholastic applications of vocabulary research, but which offended the aesthetic or ideological sensibilities of teachers who conceived of their subjects as being above the banality of "words, mere words."

In 1907, Gonzalez Lodge wrote that "the trend of thought during the last thirty years with regard to the teaching of Latin has been steadily in

the direction of greater simplification in requirements" (p. iii); he saw his Latin word-frequency count as furthering that end. The Modern Foreign Language Study (1924-1928) sponsored the word counts in Spanish, French, and German—analogs of *TTWB*. It was reported that 60% of all investigations in the modern foreign-languages field between 1930 and 1937 were vocabulary or idiom studies (Tharp, Coleman, & King, 1950, p. 471). Even the fact that first the classical languages and then the modern foreign languages came under severe challenge for their very existence in the secondary-school and college curriculum did not cause an apparent disturbance of their cordial relationship with word-counters—although without total unanimity (Henmon & Stagner, 1931). If anything, awareness of the high mortality rates in language study, and the knowledge that progressively more students (85% in 1929) limited their enrollment to two or fewer years, increased the sense that there was no time available to attempt the nonessential; it was recognized, by some at least, that "two years must be accepted as the maximum time actually available for the attainment of any and all aims by the majority of pupils" (Tharp et al., 1950, p. 465). A higher percentage of foreign-language students were beginners and the "general consensus in the profession [is] that during this phase of habit formation the vocabulary is to be rigorously limited" (Twaddell, 1966, pp. 73-74). If the literary classics were to be encountered at all, it would be in simplified versions.

"Words are the stock in trade of the shorthand writer," noted an author of the *Gregg Shorthand Simplified Textbook*, in a review of *TTWB-3*. "The more the shorthand teacher knows about the relative values of words, the more effectively he will be able to direct the shorthand learning of his pupils" (Leslie, 1946, p. 527). Given that the earliest and most famous of the 19th-century word counts were prepared for stenographic purposes, it should not come as a surprise that the materials prepared for American shorthand teachers used the modern word lists as soon as they became available: *TTWB* from 1921, then Horn's list, and then the Horn-Peterson (1943) "Basic Vocabulary of Business Letters." Word-frequency results were built into the practice materials, especially the lists of isolated words, and into shorthand speed tests (Leslie, 1944-1945). Moreover, the discovery of the skewed frequency distribution generated an important pedagogical issue for theorists and teachers: how to teach students to write the thousands of infrequently used words they would nonetheless be called upon to record and transcribe (Rowe, 1943; Humphrey, 1945; Leslie, 1945; West, 1973). The conclusion was that after rote memorization of the outlines or forms of the first few hundred most frequent words, there is little rationale for the further learning of forms. Thereafter students must rely upon acquired skill in constructing, on the spot, legible outlines for seldom-encountered words, running into the many thousands.

While shorthand enrollments were about 7% of all secondary-school

enrollments in 1961, over half of all ninth graders were enrolled in a typing course (West, 1969). The existence of so many students without occupational goals as typists raised the question of whether the acquisition of an appropriate vocabulary was not, for purposes of communication, more important than the level of typewriting skill (Lessenberry & Robinson, 1968). It was recommended that some unfamiliar words be built into the practice material of typing so that vocabulary enrichment could be an additional objective of typing courses. The use of various frequency word lists, for research into teaching methods and in the construction of connected materials for typing drills, had become standard practice (Green, 1932; Showel, 1974). Like other word fields, typewriting teachers had opted for meaningfulness—abandoning the assumption of touch-typing pioneers that letters and meaningless small units be used to initiate typing instruction.

In seeming contrast to the case of the word fields just mentioned, there exist curriculum areas where teachers and the authors of texts and courses of study have been suspicious and even hostile toward educational research. The fine arts fields come quickly to mind. In a search for the impact of research there, one encounters opposition to the utilitarianism with which research is associated; such defensive slogans appear as "art for art's sake," "music for music's sake," and "good literature justifies itself" (Moore, 1941; Funchess, 1949; Jones & Evans, 1951; Thorpe, 1958, p. 163; Beittel, 1960; Hendrickson, 1960; Hausman, 1963; Temp, 1964; Horner, 1965, pp. 4, 217). Literature will be used here as the example of a teaching field where, although reading is the essence of the classroom experience provided, its practitioners reportedly questioned and even denied the specific "logic" of vocabulary research.

In 1934 Thorndike doubted that literature courses could contribute to developing reading skills, given the small amount of reading they provided—perhaps a half million running words—and their failure to coordinate the difficulty of literary selections with the reading status of students. Although Luella Cole (Pressey, 1934) addressed her specific criticisms to history texts, what she had to say could easily be applied to literature chosen for classroom use, and to the probable results of existing practices:

Much of the dispersion of vocabulary in current texts seems to be due to efforts at style by the persons who write the textbooks. The writer would venture a guess that the style is largely wasted on the public school child and serves chiefly to add elements of confusion . . . as to discourage all but the most persistent. (p. 184)

Later in the 1930s, the supervisor of the New York City Schools' new High School Remedial Project singled out English teachers for their snobbery and misguided and myopic resistance to the proposition that "every teacher is a teacher of reading";

Teachers of English need to revise their charter of obligations to youth in terms of the contemporary situation. English teachers are wont to arrogate to themselves the holy mission of spreading the gospel of beauty and truth. . . . It seems to me that we have gone a bit too far in our emphasis on "appreciation." English teachers with a belletristic bias must first prove that their pupils possess a satisfactory measure of competence in reading. If they do not, the feast of belles-lettres must be denied these pupils until their basic reading skills have been sharpened. . . . Teachers of English have a responsibility to teach reading at all levels; this responsibility should not be shirked by any teachers working with curricula heavily weighted with the materials of pure literature. They must be reminded that English is essentially a tool and only secondarily a subject.

The concern of English teachers with "Creativism" in all its forms at the expense of basic instruction in reading is to be deplored.

A movement deserving of support of all is that calling for the simplification and revision of classics. Far too many teachers and supervisors hold the classics inviolate; their indignation at even sober efforts at editing is nothing more than a squeamishness which has no place in the make-up of persons more concerned with child development than with literature. (Scarlet, 1938, pp. 21, 23)

Teachers of literature, of the disposition of which Scarlet complained, could not make literature invulnerable to the many pressures on behalf of readability values. One such pressure came with the research reporting the greater popularity among students of easier literary materials and the competition for more readable books and magazines for youngsters' leisure reading time (Figurel, 1942). "Boys will wait for an hour at public libraries to get books by certain present day writers whose names the cultivated man has never even heard," Thorndike reminded educators in 1921. "It is commonly assumed that children and adults prefer trashy stories in large measure because they are more exciting and more stimulating in respect to sex. There is, however, reason to believe that greater ease of reading in respect to vocabulary, construction, and facts, is a very important cause of preference" (pp. 369-370). Thorndike's recommendation that "best-sellers" actually be examined for their readability not only was followed but, more important, a host of lists and services appeared which advised parents, teachers, and librarians about books of an appropriate difficulty (Vogel & Washburne, 1928; Washburne & Morphet, 1938; Morphet & Washburne, 1942; Strang, Checkovitz, Gilbert, & Scoggin, 1944; De Angelo, 1968). This, too, was known by teachers of literature.

The anthologies which appeared during the 1920s contained the beginnings of what often became elaborate study guides and glossaries—that part of the modern anthology that Lynch and Evans (1963) label the "machinery." In 1950 Gwynn wrote that "the contents of a textbook in literature are essentially stable, and literary selections tend to adhere to classical types." Yet a comparison of different editions of the "Literature

and Life" series—published in the early 1920s, revised in the mid-1930s and again in the late 1940s—led to the conclusion that the revisions "display many features indicative of attention to movements in educational thought" (pp. 242-244). The innovations of the first revision included concessions to objective testing, graded vocabulary studies, and drill.

In a satirical commentary on "Empiricus" and similar fools who would apply readability formulas or other "science" to literature, Bertrand Evans and James Lynch—both university professors of English—drew a distinction between reading materials and literature (1960). That Empiricus had done some mischief in blurring that distinction was clear in Lynch and Evans's (1963) own study of 72 high-school anthologies. They found, for example, that about one in five of the selections comprising the anthologies had been altered for high-school use, 26% of the selections in the case of 10th-grade textbooks. They were annoyed to discover 64 changes made in an extract from *Great Expectations*:

Dickens' own words, *some of which* rank as sixth-grade words on the Thorndike lists, are *sometimes* replaced by what Thorndike classed as third-grade words, although the volume is intended for the ninth-grade. (p. 51, emphasis added)

While readability research may have made only modest gains among literature teachers, these authors thought it had prospered lamentably well in the textbook publishing houses.

Sociological factors of students, subject matter, and teachers. This is an omnibus category of explanations, all of which have something, however, to do with relative status or security of place in the social structure of the schools. The thesis offered is simply that a level of schooling, or part of the whole curriculum, will be more or less receptive to, or ready for, educational innovation on the basis of social dynamics generated by status considerations. A few examples should clarify what is meant here—and reduce the "thesis" to the level of commonplace observations, familiar "truths" at least to the "old hands" who work in schools.

Students bear characteristics that, like the characteristics of teachers and of subject-matter fields, help to determine the prestige and social standing attached to teaching within a given sector of the complex organization that is the school. *In general*, it has been and remains more prestigious to be teaching most things to older students than to younger, to boys than to girls, to the academically quick than to the slow, to advanced than to beginning students, to those of socially privileged rather than of deprived backgrounds, to the scholastically successful than to the academic failures, to the white than to the nonwhite, to the academically motivated than to the vocationally oriented, to the "normal" than to the "handicapped" or "disturbed"—and to selected students rather than to all youngsters.

Curriculum fields that enroll all students, for instruction in introduc-

tory (elementary) level basic skills and knowledge, and that are taught by teachers with the most general academic credentials will be more susceptible to new theories and practices in pedagogy than other combinations of student, subject-matter, and teacher characteristics. This will be as true of nonpublic as of public schools—the engagement in vocabulary research by Catholic educators of the swelling parochial-school system was notable and long standing, reflected in numerous graduate school researches and reported in both Catholic and secular educational journals (Sister Columba, 1926; Foran, 1930; Sisters Dorothy & Cecile, 1946; Andrews, 1952; Sister Agnes, 1966). In what used to be known as the “common branches,” there is relatively less to be lost in opting for the uncertainty of change; there are relatively fewer privileges to protect. Lacking high scholastic status, such fields and their teachers may seek a second identity, perhaps the reputation of being pioneering, venturesome, creative, forward-looking.⁶

When a subject is required of all students and is preparatory to all that comes later in the curriculum, a reputation for its being unpopular with students or teachers or for producing low achievement cannot be entirely ignored. Even if teaching universal and elementary subjects does not present more difficulties, these difficulties are at least more “public” and visible. This visibility can heighten a sense of dissatisfaction with the status quo and lead to noticeable and rather rapid change, as happened in spelling and elementary-school arithmetic (Garrison & Garrison, 1929; Georges, 1929; Wise, 1934; Gorman, 1938; Buswell, 1960; Cronbach & Suppes, 1969, pp. 103–110). And in adversity there can be opportunity—particularly if optimism is a rather common personality characteristic represented in the teacher population and there is reason to believe that “solutions” are at hand.

That the subjects and teachers of the elementary school were earlier and more accepting converts to elements of progressive education than was true in secondary schools is well known. The same was true of the response to vocabulary and readability concerns, despite variations from subject to subject. The term “remedial reader” first appeared around the time of World War I and became general usage during the 1920s (Smith, 1965). As early as 1929 a private-school headmaster-author revised and simplified his secondary-school English grammar text in the realization of the “decrease in the average ability among secondary-school pupils in recent decades” (Gwynn, 1950, p. 242). In 1932 William Gray could cite research that undermined the typical assumption that all pupils entering

⁶ The issue, however, is one of low relative institutional status, not one of teacher morale. Studies consistently show higher reported job satisfaction for elementary- as compared to secondary-school teachers—a difference that cannot be explained away by the fact that women, who tend to report higher job satisfaction than do men teachers, predominate in elementary-school faculties (Clifford, 1975).

secondary schools were able readers: Studies of the reading competence "of junior and senior high-school pupils have revealed the surprising fact that many of them read no better than is normally expected of fourth- and fifth-grade pupils" (p. 29). It continued to be reported, however, that high school teachers assumed their pupils knew the words they encountered and underestimated or ignored the extent of vocabulary difficulties posed by the curriculum and textbooks (Pressey, 1934, p. 166; F. D. Curtis, 1938; Friedman, 1942; Morrison, 1962; Smith, 1965). Despite the appearance of the remedial-reading specialist, a growing body of programs and texts on treating scholastic retardation (Donnelly, 1938; Shapiro, 1938; Traxler, 1938; Brinkman, 1947) and the creation of special lists of books of known readability level for struggling secondary-school students (Carpenter, 1942), high school reading got rather small attention until the post World War II period; thereafter, "developmental reading" as distinct from remedial reading began to take on some of the aspects of conventional wisdom.

However slowly status-conscious and subject-matter-oriented high schools responded, they eventually had to deal with the consequences of becoming institutions of universal instruction. In 1930, half the youngsters aged 14 to 16 were in high school; by 1940 the proportion was two-thirds. Some relief was afforded by high dropout rates after age 16—a figure that has hovered around 25% over the past 25 years—and by curriculum and ability-tracking arrangements. A former dictionary editor (Knott, 1940), turned professor and supporter of applying vocabulary research, recalled the old days when the student who was not "book-minded" was "frozen-out, as early as possible"; by 1940 the trend was to promote such pupils until they entered trade or commercial courses or dropped out of school to work. "In the meantime, we reduce the amount of book and language activity in the school to the minimum at which we can keep our self-respect" (p. 63). That the science teachers of Connecticut, at least, continued to keep high schools "bookish places" was suggested by the results of interviews with biology teachers: About half regularly followed the vocabulary exercises, review questions, and outlines and summaries offered by their textbooks, while 11% used the suggested visual aids and but 3% took their students on recommended field trips (St. Lawrence, 1952).

Nonetheless, the message was becoming clearer. Social studies teachers were told that over 20% of secondary-school students were slow learners (Association of Social Studies Teachers of the City of New York, 1951, p. 63). Science teachers were reminded that "half the students in any class are [by definition] below average [and] . . . it may be well to select vocabulary that is from one to one and one-half grade levels below that of the grade in which the book may be used" (Mallinson, 1952, p. 274). It was reported that 10% to 15% of history textbooks was made up of "learning

activities" (Cooper, 1950). Looking back over several decades, a teachers' handbook of the Association of Social Studies Teachers of the City of New York (1951) warned:

The transformation of the high school from an institution which selected its student body and retained only those who were ready, willing, and able to measure up to predetermined standards, to one which must accept and provide for "all the children of all the people," has necessitated the re-examination of time-honored procedures. (p. 63)

A second type of status issue, "status anxiety," was manifested by school geography. For nearly a century a standard, if dreary, part of the common school curriculum, geography in the 20th century was so beset by a succession of distracting "identity problems" that one marvels that the National Council of Geography Teachers paid *any* attention to vocabulary research.

Geography was first pressured by the larger efficiency movement of the early 20th century to purge itself of the useless and inappropriate. In this case there was much to eliminate—school geography in 1900 still being what the reformers ridiculed as "sailor geography," whose "principal aim . . . seems to have been to teach the pupils the location of thousands of islands, bays, capes, peninsulas" (Whitbeck, 1921, p. 121). R. E. Dodge (1905), a leading geography educator, readily acknowledged that courses of study contained much of little value except to advanced specialists and that school geographies still resembled "compendes of great encyclopedias, which included all that was known about all phases of geography" (p. 33). That the process of elimination was not altogether successful was indicated by a high school teacher writing, in 1938, of why geography teaching was ineffective and why teachers showed such apathy toward the subject: "Authors . . . have contributed . . . bulky geography texts that not only present a forbidding appearance but also contain a verbose, technical content much of which might have been omitted" (Levi, 1938, p. 187).

There was no denying that reading difficulties plagued this traditional bookish teaching. "Ask an elementary teacher about her problems in teaching geography and you are almost sure to find that she considers reading the word matter, or textual material of the textbook, her major problem," a Pennsylvania teacher stated in a much-cited article (Schauer, 1945, p. 285). After an early study of a geography text—using *TTWB-1* and encouraged by Buckingham—there appeared occasional research, reviews, and pedagogical discussion that stressed vocabulary reform, a goodly portion of it coming from teachers (Ware, 1921; Gabler, 1923; Upton, 1933; Wilson, 1937; D. Curtis, 1938; Zink, 1938; Anderzhon, 1950; Harbeson, 1951; Elkins, 1956; Phillips, 1956; Chall, 1958a, pp. 125–130; Thralls, 1958; James & Crape, 1968).

In place of geography as "nothing but compilations of miscellaneous

geographical information coupled with an inordinate amount of locational geography" (Whitbeck, 1921, p. 121), geographers sought to organize geographical knowledge and concepts around one or another focus, the succession preoccupying curriculum makers and teachers through much of the 1920s and 1930s (Kolevzon & Maloff, 1964). Granting "home geography" in the primary grades, what thereafter and for how long: human geography? physical geography? cultural geography? economic geography? regional geography? one-world geography? resource geography? commercial geography? And on the "one-cycle" or the "two-cycle" plan?

Before these issues reached resolution, geography faced the ultimate threat to its status: the pressure that it be merged with history, the study of current events, and bits of the other social sciences into "social studies." Such a possibility of "fusion" commanded most of school geography's attention through the 1930s. Along with facing a loss of its identity as a separate school subject, geography contemplated the likelihood of total extinction. Geography had entered the college curriculum late and never achieved the status of a generally required subject. A lack of geographical competence was often blamed for the condition of geography books and for the reliance on memoriter teaching methods (Whitbeck, 1921). Given the continuing lack of teacher preparation in geography and an accompanying dislike of the subject, it was legitimate to fear that geography would eventually disappear even from "the new social studies."

The textbook committee of the National Society for the Study of Education had noted in 1931 that "as one passes from the fields of spelling and arithmetic to the social studies . . . one notices a striking contrast [of] comparatively few well-planned investigations" (Schorling & Edmonson, 1931, p. 51). It was seen that the comparatively few investigations of geography teaching came rather late and were concentrated in unpublished studies (Zink, 1938; McAulay, 1948). Geography shared with history teaching an apparent reluctance to conceive of its pedagogical difficulties as heavily influenced by language problems. Moreover, the new social studies were busied, more so than other fields, with trying out the variety of *nonreading* activities advocated by progressive theorists: field trips, construction projects, dramatic play, art activities, radio and film use.

Given the pressures facing school geography, it is not surprising to read that a 10-year span existed between the articulation of a new pedagogical truth by geography authorities and its incorporation into the objectives, content, and methodology of textbooks in this field; another 10 years elapsed before its appearance in courses of study (McAulay, 1951). It is generally admitted that textbooks in geography and the other social studies eventually improved (Carpenter, 1954). Yoakam's (1945) readability formula was applied to geographies published in the early 1940s and showed the clear trend toward simpler materials (p. 307). An editor of a

major publishing house assured a large reading conference that publishers knew of the demand for easy social studies books and were making "sincere and tireless" efforts to produce them (Anderson, 1950, p. 90). High school history teachers perceived progress in textbook readability and asked for more (Gross, 1952). Geography and other social studies material were somehow controlled by Yoakam's readability formulas for *My Weekly Reader* (Yoakam, 1945, p. 305). School encyclopedias and other reference books became much more appropriate for their intended audiences—an accomplishment owed to the vocabulary researchers who began to publicize their findings (Overholser, 1935; Jensen, 1939; Edgerton, 1945, 1954).

Algebra is our last illustration, representing speculation on a third kind of status consideration influencing the acceptance of research into school practice. In this case, it is algebra teachers' status as *mathematics* teachers that appears to be a factor limiting the influence that comes through *educational* research.⁷

Thorndike's *The Psychology of Algebra* was published in 1923. Its recommendations showed predictable congruity with his other educational writings, a consistency imposed by his overarching learning theory. As in his influential works on arithmetic and vocabulary, Thorndike emphasized attention to the formation of the desired stimulus-response bonds and the preclusion of learning bad habits, the provision for adequately distributed and sufficient practice, restriction of the curriculum to the most useful topics and important skills, and the elimination of improbable and unnecessarily abstract problems.

This book met the requirement of timeliness, for a reform movement in algebra teaching was already under way. In his *Algebra*, Thorndike referred to the National Committee on Mathematical Requirements which had been formed in 1916 by the Mathematical Association of America. The committee's charge was to make recommendations for reorganizing courses and improving teaching of mathematics at the secondary and collegiate levels (Young, 1921). The committee began with a membership of six mathematicians and three teachers of high school mathematics. The results, to which many teachers and teacher organizations contributed, were reported in the publication of the National Council of Teachers of Mathematics, *The Mathematics Teacher*. The conclusions about the "new algebra" were in line with the "new psychology" in rejecting mental discipline as a justification for mathematics education and in accepting modern theory on drill and transfer of training. Textbooks subsequently published reflected the committee's recommendations.

Despite understandable differences in emphasis between Thorndike's

⁷ William Rupley has begun an investigation of Thorndike's work on algebra in relation to his and others' efforts to reform the teaching of algebra. I am indebted to Mr. Rupley for access to his preliminary research and for conversations that have informed my speculations here.

Algebra and the National Committee's report, there was considerable overlap and agreement. That the community of mathematics teachers was aware of Thorndike's *Algebra* shows in the early articles on the algebra debated in *The Mathematics Teacher*, although the National Committee report got more mentions. When the National Council of Teachers of Mathematics published *The Teaching of Algebra* in 1932, however, Thorndike's work was unmentioned; its authors were mathematics educators, not educational researchers or psychologists. Unlike most teachers of elementary-school mathematics, teachers of algebra are specialists in mathematics. Although firm evidence is not yet at hand, it is at least probable that algebra teachers highly value their status as experts in mathematics and not as "educationists." Their average relative levels of commitment to the discipline of mathematics as compared to a learner-centered psychology can probably be determined by further investigation.

Conservatism in the teaching of algebra may well have been fostered by the restrictions on "outside influence" imposed by the self-image of the mathematics teacher. Unquestionable conservatism is encouraged, to date, by the isolation of algebra from the pressures of mass education. Except for the college-bound high-school population—until recently a small minority of even high-school graduates—prospects for failure in traditional algebra can be, and have been, counseled into general or business mathematics or advised to drop the subject if they encountered insurmountable difficulties. Thus algebra teachers have been denied the opportunity to acquire considerable first-hand experience on the question of the contribution of language to the difficulties of their subject matter. Also thus deprived have been university professors of mathematics—despite their complaints that most college students are mathematical morons. It should not be surprising, then, that the innovative mathematics curricula to which university mathematicians contributed so heavily during the post-Sputnik years were found to require more reading and present more reading difficulties than more traditional mathematics materials; parts of Grade 4 materials had Grade 10 reading matter. The investigators (Smith & Heddens, 1964) concluded: "It would appear that the merits of the discovery method are largely immaterial if Johnny cannot read his book."

A cumulative research tradition. In the beginning of this chapter, the reader was asked to consider whether vocabulary and readability research displayed the cumulative properties which are believed to enhance the potential of research to affect everyday behavior. Observing the history of this research, not as an expert in language investigations but as a historian, vocabulary and readability research seems to have passed this test rather well. How far short this particular tradition of research falls of designation as a "paradigm," in the way Kuhn (1962) conceives the history

of science, is a discussion for another time and place. Suffice it to record that there is an impressive continuity from the earliest, least mathematical word counts to complex (if not necessarily elegant) and linguistically aware readability formulas—a continuity that appears to represent the processes of refinement, integration, correction, and elaboration that “cumulation” entails in its more positive connotations.

In a survey for the most recent *Encyclopedia of Educational Research*, MacGinitie (1969) commented that vocabulary lies somewhat outside current lines of investigation into language development, “but in many practical respects remains an index of development outweighing all others” (p. 693). As both he and Klare (1963) have also observed, a vocabulary factor of some sort has remained the one stable element in the numerous measures of readability which extend from Lively and Pressey in 1923 and Vogel and Washburne in 1928 to Taylor’s (1953) “Cloze test” and Bormuth’s (1966) formulation—one of whose many variables is the word-frequency index of *TTWB*-3. Factors included in readability formulas might be quantitative. The vocabulary variable might be “number of different words” or “percent of hard words” as determined by comparison with a frequency list like *TTWB* or a “familiarity” list like Edgar Dale’s. Or it might include qualitative variables like Ojemann’s (1934) “obscurity” factor. Other variables—such as dimensions of style, shown perhaps by “number of simple sentences” or “percentage of prepositional phrases,” and dimensions of interest, shown perhaps by “percentage of colorful words” or “human interest words”—could be dropped from a formula if they proved to add more to difficulty than they added to the prediction of readability. But vocabulary could not be eliminated from a readability formula, having proved itself the key (though not sole) factor in reading comprehension and the one variable most related to all other criteria of difficulty so far explored (Lorge, 1944; Dale & Chall, 1948; Chall, 1958a; Klare, 1963). That the existing research corpus has generated rather precisely defined areas of still-needed measurements of other readability dimensions should also be acknowledged; such is, after all, one of the dividends of a cumulative research tradition.

The tradition of readability research represents an intricate linkage or network—not of discrete or isolated investigations and applications but of extensions and modifications of key studies. In graphic form this network would resemble more a family tree of research than a heap of studies. *In large measure, behind this conceptual and empirical network lies a personal-historical network: a system of communication of interests between teachers and their students, an “old-boy” set of academic and research relationships clustering around a few universities but radiating influence into publishing centers as well as to other institutions. An initial attempt at representing this web of personal relationships in vocabulary and read-*

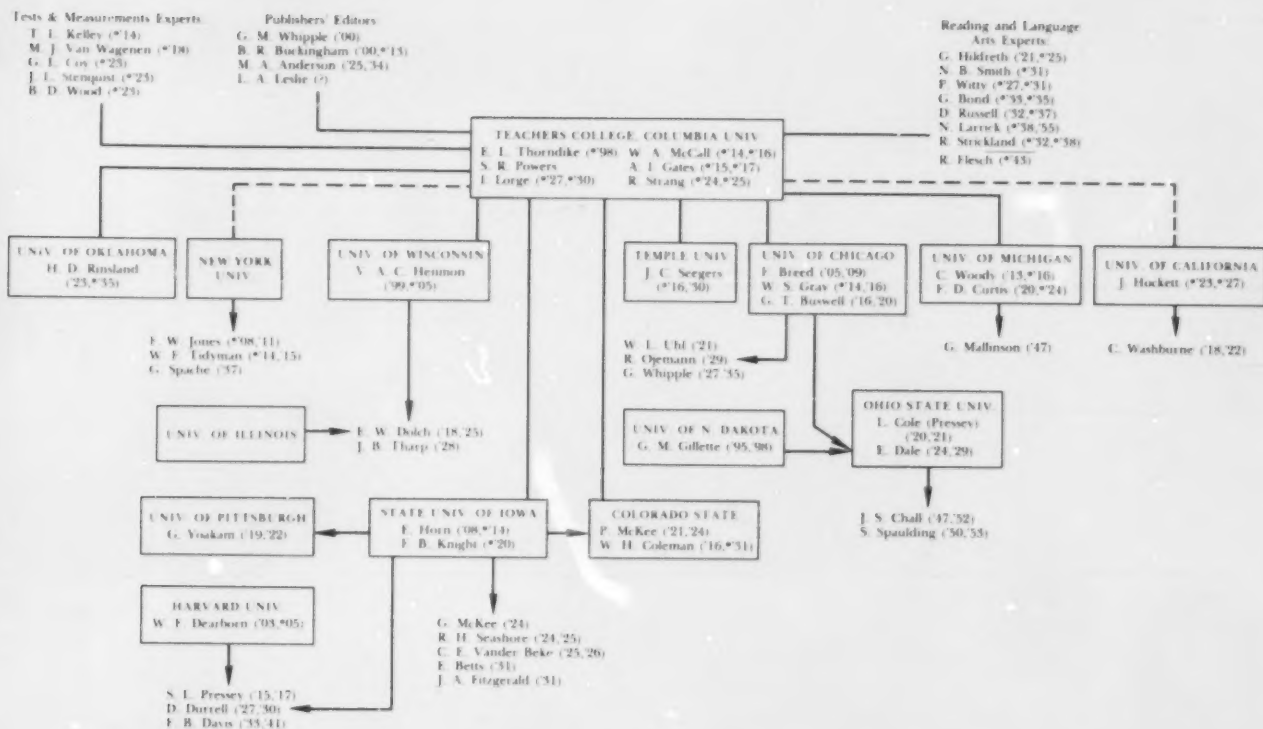


FIGURE 2. The personal network in vocabulary research: A preliminary outline. (Asterisk indicates Columbia University graduate degree and date.)

ability research is shown in Figure 2. Fuller biographical investigation would undoubtedly identify and clarify the participation of a few additional centers, such as Colorado State College of Education at Greeley where W. H. Coleman (PhD, Teachers College) and Paul McKee (PhD, Iowa) taught and wrote about vocabulary, spelling, and reading. Further elaboration would, however, very likely only intensify our awareness of the massive centrality and insistent outreach of E. L. Thorndike and Teachers College, Ernest Horn and the State University of Iowa, and W. S. Gray in reading and Guy Buswell in arithmetic vocabulary at the University of Chicago. Thorndike's crucial position only grows more striking when one remembers that Horn and Gray came under Thorndike's influence at Teachers College.

During Thorndike's years at Teachers College, that institution surpassed all others in producing leaders for the world of elementary and secondary education (Joncich, 1968). Many of the prospective teachers, supervisors, superintendents, health educators, curriculum specialists, and counselors studying at Teachers College never worked with, or even saw, Thorndike—especially after 1921 when the Institute of Educational Research was established and his teaching contacts were lessened. Nonetheless, his very presence at the institution and the omnipresence of *TTWB* as a research tool caused his influence to pass through closed doors.

Thorndike's presence also opened doors. The author of *Adult Learning* in 1928 and *Adult Interests* in 1935, Thorndike was credited with having done a great deal to stimulate research and educational programs for adults. He was honored by the American Association for Adult Education by being chosen its president in 1935 (Lorge, 1940; Joncich, 1968). At the same time, this Association made the initial grant to open a Readability Laboratory at Teachers College, under the direction of the gifted humanist Lyman Bryson. In its 10-year life the laboratory sought to integrate the knowledge on readability from several disciplines and from the publishing field and to add to the knowledge of the less easily measured aspects of readability (Bryson, 1936; Chall, 1958a). An experimental book series, the "People's Library," was produced as an attempt to provide better nonfiction reading material for average adult readers.

It was at the Readability Laboratory in the early 1940s that the expatriate from Vienna, Rudolph Flesch (PhD, 1943), became interested in readability. The maverick Flesch added formulas which entered the cumulative readability tradition. And the popular success of Flesch's *Why Johnny Can't Read* (1955) made the pedagogy of reading a political issue and reopened intense controversy about the initial teaching of reading, controversy which had been largely muted during the preceding half century.

Had he been alive in 1955, Thorndike might well have expressed wry amusement, both at the instant conversion of public opinion to phonics

and at the sometimes hysterical and obscurantist reaction of the professional reading expert—seeing, perhaps, in both sets of responses some evidence of “underprivilege in respect to intellect.” But, then, Thorndike was never as interested in the acquisition of initial skill in reading as he was in improving the possibilities for long-term growth in reading.

Timeliness. Finally, and in conclusion, one of the functions of educational research which has been much too little considered and appreciated is that of legitimation: giving the support of facts and the prestige of science to existing trends and movements (Clifford, 1975, pp. 27–28). Vocabulary research nicely complemented the teaching and curriculum reforms advocated by the visionaries of the early progressive education movement. Supporters of the “new education” received welcomed help from the vocabularists in their own struggles to “vitalize” the curriculum, emphasize the meaningful in classroom experiences, recognize and provide for group and individual differences, and equip students for a socially conscious adulthood.

In opposing the “lock-step” pedagogy of the “old education,” the progressives encouraged teachers to create “centers of interest” and units of study, in which students could work as individuals or in small groups with a variety of materials. Information about the readability of printed material and about students’ language experiences at various ages and stages of development were presumed to equip teachers to make intelligent choices within a pedagogical system that gave them more decisions to make. (One of the purposes of Hockett’s much-cited study in 1938 of the readers used in California schools was to enhance teachers’ abilities to function in precisely these ways.)

When Thorndike (1934, p. 3) ridiculed the wastefulness of a pupil’s being required to memorize the definition of a bay—“a body of water partly surrounded by land”—when he does not know the meaning of *body*, *partly*, or *surrounded*, he was responsive, as well as appealing, to a corpus of administrative and public opinion that was already highly sensitive to issues of waste and inefficiency. Standardization was one result of the quest to root out the causes of inefficiency and irrationality in the educational system. The greater uniformity of textbooks, in content and vocabulary, as a result of textbook analyses fits this preexisting movement as well.

The progressives added a child-centeredness to the content of the reading curriculum and a desire that schooling be as pleasurable and failure-free as possible, while vocabulary research intensified and made more objective the existing tendencies toward simplification that supported the same goals. No progressive could disagree with the humane and commonsensical import of what Thorndike (1934) had to say about the need to improve school reading materials:

The worst, which often occurs in basic readers and supplementary reading, is to require children to read material which they cannot understand; or can understand, but do not enjoy; or cannot understand, and would not enjoy if they could understand it. . . . Books do not have to be bad literature; but the vocabulary and sentence structure must not thwart comprehension of what the book tells and that must be something the pupil cares to be told. (p. 19)

Thorndike's debunking of the useless and recondite in traditional textbooks was more than an attack upon uncommon and unimportant words; it was part of a much larger campaign against a false psychology of mental discipline, against the scholar's arrogances and the teacher's self-deceptions, and on behalf of a clear specification of the social utilities of schooling and a determination to fit teaching procedures more precisely to these objectives.

REFERENCES

- Agnes, Sister M. How readable are Primary religion books? *Catholic School Journal*, 1966, **66**, 60-61.
- Anderson, M. A. Learning to read in the social studies. In G. A. Yoakam (Ed.), *Reading and the language arts in the school program*. (A Report of the Sixth Annual Conference on Reading.) Pittsburgh: University of Pittsburgh Press, 1950.
- Anderson, M. A. The values and uses of vocabulary control. In J. A. Figurel (Ed.), *Reading for effective living*. (International Reading Association Conference Proceedings, Vol. 3.) New York: Scholastic Magazines, 1958.
- Anderzhon, M. L. Review of *Cross-country geography for children* by Hanna and Kohn. *Journal of Geography*, 1950, **49**, 304.
- Andrews, D. I. A study of the vocabulary and readability of a third-grade classroom periodical. *National Catholic Education Association Bulletin*, 1952, **49**, 221-225.
- Association of Social Studies Teachers of the City of New York. *A handbook for social studies teaching*. New York: Republic Book Company, 1951.
- Ayer, A. M. Some difficulties in elementary school history. *Teachers College Contributions to Education*, 1926, **212**.
- Ayer, F. C. *A study of high school spelling vocabulary*. Austin, Tex.: Steck, 1945.
- Ayres, L. P. *Laggards in our schools*. New York: Charities Publication Committee, 1909.
- Ayres, L. P. *A measuring scale for ability in spelling*. New York: Russell-Sage Foundation, 1913. (a)
- Ayres, L. P. *The spelling vocabularies of personal and business letters*. New York: Russell-Sage Foundation, 1913. (b)
- Ayres, L. P. History and present status of education measurements. In G. M. Whipple (Ed.), *The measurement of educational products*. (Seventeenth Yearbook of the National Society for the Study of Education, Part 2.) Bloomington, Ill.: Public School Publishing Company, 1918.
- Babbitt, E. H. A vocabulary test. *Popular Science Monthly*, 1907, **70**, 378.
- Bachmann, H. M. A semantic study of the books of two authors dealing with classical antiquity. *University of Pittsburgh Bulletin*, 1944, **40**, 16-27.
- Baker, F. M. A tentative vocabulary for first year French students. *Journal of Educational Research*, 1928, **18**, 369-377.

- Barnhart, C. L. Contributions of Dr. Thorndike to lexicography. *Teachers College Record*, 1949, **51**, 35-42.
- Barton, A. H., & Wilder, D. E. Research and practice in the teaching of reading: A progress report. In M. B. Miles (Ed.), *Innovation in education*. New York: Teachers College, 1964.
- Beittel, K. R. Art. In C. W. Harris (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1960.
- Bellugi, U. Review of *Language in the crib* by R. Weir. *Harvard Educational Review*, 1964, **34**, 332-334.
- Bennett, H. E. *School efficiency: A manual of modern school management*. Boston: Ginn, 1917.
- Berry, A. Development of reading vocabulary and word recognition. 1. *Reading in the elementary school*. (Forty-eighth Yearbook of the National Society for the Study of Education, Part 2.) Chicago: University of Chicago Press, 1949.
- Betts, E. A. Readability: Its application to the elementary school. *Journal of Educational Research*, 1949, **42**, 438-459. (a)
- Betts, E. A. *Second spelling vocabulary study—Grade placement of words in eight recent spellers*. New York: American Book Company, 1949. (b)
- Beust, E. The publisher's role in promoting growth in interpretation. In W. S. Gray (Ed.), *Promoting growth toward maturity in interpreting what is read* (Supplementary Educational Monograph, No. 74). Chicago: University of Chicago Press, 1951.
- Bloomfield, L. *Language*. New York: Holt, Rinehart, & Winston, 1933.
- Boas, G. *The cult of childhood*. London: University of London, 1966.
- Bolton, F. B. A study of vocabulary in the social studies. *Social Education*, 1943, **7**, 17-18.
- Bormuth, J. R. Readability: A new approach. *Reading Research Quarterly*, 1966, **1**, 79-132.
- Brinkman, A. P. Reading improvement in geography for retarded readers. *Journal of Geography*, 1947, **46**, 234-240.
- Brown, R. Vocabularies of history and reading textbooks. *Bulletin of the Department of Elementary School Principals*, 1931, **10**, 408-411.
- Brown, R. M. Fifth-grade geography textbooks: A criticism. *Journal of Geography*, 1920, **19**, 172-185.
- Bryson, L. Readability laboratory. *Library Journal*, 1936, **61**, 455.
- Buchanan, M. A. *A graded Spanish word book*. Toronto: University of Toronto Press, 1927.
- Buckingham, B. R. Spelling ability, its measurement and distribution. *Teachers College Contributions to Education*, 1913, **59**.
- Buckingham, B. R. Textbooks. In C. W. Harris (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1960.
- Buckingham, B. R., & Dolch, E. W. *A combined word list*. Boston: Ginn, 1936.
- Buros, O. *Mental measurement yearbook* (4 vols.). Highland Park, N. J.: Gryphon Press, 1941-1968.
- Buswell, G. T. Psychology of learning: General conditions affecting teaching and learning. *Review of Educational Research*, 1933, **3**, 338-348.
- Buswell, G. T. Arithmetic. In C. W. Harris (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1960.
- Buswell, G. T., & John, L. *The vocabulary of arithmetic*. Chicago: University of Chicago Press, 1931.
- Buswell, G. T., & Judd, C. H. *Summary of educational investigations relating to arithmetic*. Chicago: University of Chicago Press, 1925.

- Carpenter, H. M. *Gateways to American history: An annotated graded list of books for slow learners in junior high school*. New York: Wilson, 1942.
- Carpenter, H. M. (Ed.). *Skills in social studies*. (1953 Yearbook of the National Council for the Social Studies.) Washington, D.C.: The National Council for the Social Studies, 1954.
- Carroll, J. B., Davies, P., & Richman, B. *The American Heritage word frequency book*. New York: American Heritage, 1971.
- Carroll, J. B., & White, M. N. Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 563-576.
- Chall, J. S. A survey of users of the Dale-Chall formula. *Educational Research Bulletin*, 1956, **35**, 197-212.
- Chall, J. S. *Readability: An appraisal of research and application*. Columbus: Ohio State University, 1958. (a)
- Chall, J. S. What about controlled vocabulary? The history of controlled vocabulary. In J. A. Figurel (Ed.), *Reading for effective living*. (International Reading Association Conference Proceedings, Vol. 3.) New York: Scholastic Magazine, 1958. (b)
- Chall, J. S. *Learning to read: The great debate*. New York: McGraw-Hill, 1967.
- Chambers, W. G. How words get meaning. *Pedagogical Seminary*, 1904, **11**, 30-50.
- Charters, W. W. *Teaching the common branches: A textbook for teachers* (Rev. ed.). Boston: Houghton Mifflin, 1924.
- Chase, S. E. Waste in arithmetic. *Teachers College Record*, 1917, **18**, 360-370.
- Chauncey, H., & Dobbin, J. E. *Testing: Its place in education today*. New York: Harper & Row, 1963.
- Chester, R., & Otto, W. Children's sight words—Printed materials or oral production. *Journal of Educational Research*, 1974, **67**, 247-252.
- Clark, B. R. *Educating the expert society*. San Francisco: Chandler, 1962.
- Clement, J. A., & Dolch, E. W. Textbooks. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1941.
- Clifford, G. J. A history of the impact of research on teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand-McNally, 1973.
- Clifford, G. J. *The shape of American education*. Englewood Cliffs, N. J.: Prentice-Hall, 1975.
- Cole, L. *Psychology of the elementary school subjects*. New York: Farrar & Rinehart, 1934.
- Cole, L. *The teacher's handbook of technical vocabulary*. Bloomington, Ill.: Public School Publishing Company, 1940.
- Cole, R. D. *Modern foreign languages and their teaching* (Rev. ed. by J. B. Tharp). New York: Appleton-Century-Crofts, 1937.
- Coleman, A. *The teaching of modern foreign languages in the United States: A report prepared for the Modern Foreign Language Study*. New York: Macmillan, 1929.
- Coleman, A. The curriculum: Modern foreign languages. *Review of Educational Research*, 1934, **4**, 156-160.
- Coleman, A. *An analytical bibliography of modern language teaching, 1932-1937* (Vol. 2). Chicago: University of Chicago Press, 1938.
- Coleman, A., & King, C. B. Foreign languages: Modern. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1941.
- Coleman, J. S. *The adolescent society*. New York: Free Press, 1961.
- Coleman, W. H. *A critique of spelling vocabulary investigation*. Greeley: Colorado State Teachers College, 1931.
- Columba, Sister M. A study of interests and their relation to other factors of

- achievement in the elementary school subjects. *Educational Research Bulletin*, 1926, 1(7).
- Colvin, C. M. A re-examination of the vocabulary question. *Elementary English*, 1951, 28, 350-356.
- Cook, W. W. Test or achievement. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1941.
- Cooper, K. S. How good are American history textbooks? *Social Education*, 1950, 14, 341-343.
- Cornell, E. A beginning reading vocabulary for foreign-born adults with special reference to those who cannot read in any language. *University of the State of New York Bulletin*, 1930, 948.
- Cremin, L. A. *The transformation of the school*. New York: Knopf, 1961.
- Crocker, L. Teaching English conversation in Japan. *Quarterly Journal of Speech*, 1945, 31, 63-66.
- Cronbach, L. J. The learning process and text specifications. In L. J. Cronbach (Ed.), *Text materials in modern education*. Urbana: University of Illinois Press, 1955.
- Cronbach, L. J., & Suppes, P. (Eds.). *Research for tomorrow's schools*. New York: Macmillan, 1969.
- Cureton, E. E. Validity. In E. F. Lindquist (Ed.), *Educational measurement*. Washington, D.C.: American Council on Education, 1951.
- Curtis, D. Reading as it functions in fifth-grade geography. *National Elementary Principals*, 1938, 17, 376-378.
- Curtis, F. D. Some values derived from extensive reading of general science. *Teachers College Contributions to Education*, 1924, 163.
- Curtis, F. D. *A digest of investigations in the teaching of science in the elementary schools*. Philadelphia: Blakiston, 1926. (a)
- Curtis, F. D. A study of the vocabulary of scientific articles appearing in daily newspapers. *School and Society*, 1926, 23, 821-824. (b)
- Curtis, F. D. *Second digest of investigations in the teaching of science*. Philadelphia: Blakiston, 1931.
- Curtis, F. D. *Investigations of vocabulary in textbooks of science for secondary schools*. Boston: Ginn, 1938.
- Curtis, F. D. *Third digest of investigations in the teaching of science*. Philadelphia: Blakiston, 1939.
- Dale, E. A comparison of two word lists. *Educational Research Bulletin*, 1931, 10, 484-489. (a)
- Dale, E. Difficulties in vocabulary research. *Educational Research Bulletin*, 1931, 10, 119-122. (b)
- Dale, E. Readability in the primary grades. In G. A. Yoakam (Ed.), *A report of the third annual conference on reading*. University of Pittsburgh. Pittsburgh: University of Pittsburgh, 1947.
- Dale, E. *Bibliography of vocabulary studies*. Columbus: Ohio State University, Bureau of Educational Research, 1949.
- Dale, E. The problem of vocabulary in reading. *Educational Research Bulletin*, 1956, 35, 113-123.
- Dale, E., & Chall, J. S. A formula for predicting readability. *Educational Research Bulletin*, 1948, 27, 11-20, 28.
- Dale, E., & Reichert, D. *Bibliography of vocabulary studies* (Rev. ed.). Columbus: Ohio State University, 1957.
- Danzger, M., & Halpern, H. Relation of stuttering to word abstraction, part of speech, word length, and word frequency. *Perceptual and Motor Skills*, 1973, 37, 959-962.

- Davis, F. B. The interpretation of frequency ratings obtained from *The teacher's word book*. *Journal of Educational Psychology*, 1944, **35**, 169-174.
- De Angelo, R. W. Media for disseminating critiques. *Perspectives in Reading*, 1968, **10**, 110-125.
- Dearborn, F. R. A study of erroneous word concepts in reading. *Elementary English Review*, 1929, **6**, 3-6, 23.
- Dewey, J. *How we think*. Boston: Heath, 1910.
- Dewey, J. C. *A case study of reading comprehension difficulties in American history*. Unpublished doctoral dissertation, State University of Iowa, 1931.
- Dickerson, D. F. Misleadings vs. actualities in spelling. *American School Board Journal*, 1950, **120**, 33-34.
- Dodge, R. E. The archaic in geography teaching. *Journal of Geography*, 1905, **4**, 32-33.
- Dolch, E. W. *Reading and word meanings*. Boston: Ginn, 1927.
- Dolch, E. W. How much word knowledge do children bring to grade one? *Elementary English Review*, 1936, **13**, 177-183.
- Dolch, E. W. The Bible is easy reading. *Elementary English Review*, 1938, **15**, 297-298.
- Dolch, E. W. *Psychology and the teaching of reading* (2nd ed.). Champaign, Ill.: Garrard Press, 1951.
- Dolch, E. W. *Better spelling*. Champaign, Ill.: Garrard Press, 1960.
- Donnelly, H. E. The remedial reading classroom. *Education*, 1938, **59**, 31-36.
- Doran, E. W. A study of vocabularies. *Pedagogical Seminary*, 1907, **14**, 401-438.
- Dorothy, Sister M., & Cecile, Sister R. A vocabulary analysis of recently published pre-primers. *Journal of Educational Research*, 1946, **40**, 116-125.
- Durrell, D., & Sullivan, H. B. Vocabulary instruction in the intermediate grades. *Elementary English Review*, 1938, **15**, 138-145, 160, 185-198.
- Edgerton, R. B. How difficult are children's encyclopedias? *Elementary School Journal*, 1945, **45**, 378-385, 455-464.
- Edgerton, R. B. How difficult are children's encyclopedias? A second report. *Elementary School Journal*, 1954, **55**, 219-225.
- Elkins, A. D. The problems of reading geography. *Education*, 1956, **77**, 37-44.
- Englehart, M. D., & Beck, J. M. The improvement of tests. In W. G. Findley (Ed.), *The impact and improvement of school testing programs*. (Sixty-second Yearbook of the National Society for the Study of Education, Part 2.) Chicago: University of Chicago Press, 1963.
- Evans, B., & Lynch, J. J. *Dialogues on the teaching of literature*. New York: Bookman, 1960.
- Evenden, E. S. Contributions of research to the education of teachers. In G. M. Whipple (Ed.), *The scientific movement in education*. (Thirty-seventh Yearbook of the National Society for the Study of Education, Part 2.) Bloomington, Ill.: Public School Publishing Company, 1938.
- Faucett, L., & Maki, I. *A study of English word-values statistically determined from the latest extensive word-counts*. Tokyo: Matsumara Sanshodo, 1932.
- Faucett, L., Palmer, H., Fife, R. H., West, M., & Thorndike, E. L. *Interim report on vocabulary selection*. London: King, 1936.
- Fennell, E. L. Recurring words and their relation to difficulties in comprehension. *Elementary School Journal*, 1928, **42**, 42-53.
- Figurel, J. A. Relative difficulty of reading material for ninth grade literature. *Pittsburgh Schools*, 1942, **16**, 125-138.
- Flesch, R. *Why Johnny can't read*. New York: Harper & Row, 1955.
- Foran, T. G. The spelling vocabulary. *Catholic Educational Review*, 1930, **28**, 587-600.

- Foran, T. G. The vocabulary of primary reading. *Catholic Educational Review*, 1934, **32**, 596-607.
- Forbes, F. W., & Cottle, W. C. A new method for determining readability of standardized tests. *Journal of Applied Psychology*, 1953, **37**, 185-190.
- Francis, W. N. The Brown University standard corpus of English: Some implications for TESOL. In B. W. Robinett (Ed.), *On teaching English to speakers of other languages*. Washington, D.C.: Teachers of English to Speakers of Other Languages, 1967.
- Franzen, R. H., & Knight, F. B. *Textbook selection*. Baltimore: Warwick & York, 1922.
- Friedman, K. C. A word test in the social studies. *Social Studies*, 1942, **33**, 268-296.
- Fries, C. C. *English word lists: A study of their adaptability for instruction*. Ann Arbor, Mich.: Wahr, 1950.
- Fries, C. C. *Linguistics and reading*. New York: Holt, Rinehart & Winston, 1963.
- Funchess, L. W. Research is needed in music education. *Phi Delta Kappan*, 1949, **30**, 349-350.
- Gabler, E. R. A geography vocabulary study. *Journal of Geography*, 1923, **22**, 104-108.
- Gansl, I. Vocabulary: Its measurement and growth. *Archives of Psychology*, March 1939, pp. 1-52.
- Garrison, S. C., & Garrison, K. C. *The psychology of elementary school subjects*. Richmond, Va.: Johnson, 1929.
- Gates, A. I. Contributions to the psychology of the elementary school subjects. *Teachers College Record*, 1926, **27**, 548-557.
- Gates, A. I. *A reading vocabulary for the primary grades*. New York: Teachers College, 1935. (Originally published, 1926.)
- Gates, A. I. Vocabulary control in basal reading materials. *Reading Teacher*, 1961, **15**, 81-85.
- Gates, A. I., Bond, G. L., & Russell, D. H. Relative meaning and pronunciation difficulties of the Thorndike 20,000 words. *Journal of Educational Research*, 1938, **32**, 161-167.
- Georges, J. S. The nature of difficulties encountered in reading mathematics. *School Review*, 1929, **37**, 217-226.
- Gillette, J. M. Extent of personal vocabularies and cultural control. *Scientific Monthly*, 1929, **29**, 451-457.
- Gorman, F. H. The arithmetic vocabulary of the elementary school teacher. *Elementary School Journal*, 1938, **38**, 373-379.
- Gray, W. S. *Summary of investigations relating to reading* (Supplementary Educational Monographs, No. 28). Chicago: University of Chicago, 1925.
- Gray, W. S. Summary of reading investigations. *Elementary School Journal*, 1929, **29**, 413-457, 496-509.
- Gray, W. S. Special methods on the high-school level: Reading and literature. *Review of Educational Research*, 1932, **2**, 29-34, 86-87.
- Gray, W. S. Progress in the study of readability. In *Library trends*. Chicago: University of Chicago Press, 1937.
- Gray, W. S. Contributions of research to special methods: Reading. In G. M. Whipple (Ed.), *The scientific movement in education*. (Thirty-seventh Yearbook of the National Society for the Study of Education, Part 2.) Bloomington, Ill.: Public School Publishing Company, 1938.
- Gray, W. S., & Leary, B. E. *What makes a book readable?* Chicago: University of Chicago Press, 1935.
- Green, H. H. The relative effectiveness of the thousand commonest words in the teaching of typewriting. *University of Iowa Studies in Education*, 1932, **5**, 167-178.

- Greene, H. A. English—Language, grammar, and composition. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1941.
- Greene, H. A. English—Language, grammar and composition. In W. S. Monroe (Ed.), *Encyclopedia of educational research* (Rev. ed.). New York: Macmillan, 1950.
- Greene, H. A., Jorgensen, A. N., & Gerberich, J. R. *Measurement and evaluation in the elementary school*. New York: Longmans, Green, 1953.
- Gregory, C. A. The reading vocabularies of third-grade children. *Journal of Educational Research*, 1923, **7**, 127-131.
- Grinstead, W. J. On the sources of the English vocabulary: A report on the Latin-English word count of the American Classical League, Foreword by Gonzalez Lodge. *Teachers College Record*, 1924, **26**, 32-46.
- Gross, R. E. American history teachers look at the book. *Phi Delta Kappan*, 1952, **33**, 290-291.
- Guilford, J. P., & Zimmerman, W. S. The Guilford-Zimmerman aptitude survey. *Journal of Applied Psychology*, 1948, **32**, 24-32.
- Gwynn, J. M. *Curriculum principles and social trends* (Rev. ed.). New York: Macmillan, 1950.
- Haggerty, M. E. Specific uses of measurement in the solution of school problems. In G. M. Whipple (Ed.), *The measurement of educational products*. (Seventeenth Yearbook of the National Society for the Study of Education, Part 2.) Bloomington, Ill.: Public School Publishing Company, 1918.
- Hall, G. S. The contents of children's minds. *Princeton Review*, 1883, **11**, 249-272.
- Hall-Quest, A. L. *The textbook: How to use it and judge it*. New York: Macmillan, 1918.
- Hamilton, M. H. *Studies in the vocabulary of second year French reading texts*. Unpublished master's thesis, University of Iowa, 1932.
- Hanna, P. R., Sabaroff, R. E., Davies, G. F., & Farrar, C. R. *Geography in the teaching of social studies, concepts and skills*. Boston: Houghton Mifflin, 1966.
- Harap, H. The curriculum, evaluation of courses of study and textbooks. *Review of Educational Research*, 1934, **4**, 196-197, 242-244.
- Harbeson, C. Intermediate geography: A reading problem. *Journal of Geography*, 1951, **50**, 21-24.
- Hartmann, G. W. Further evidence on the unexpected large size of recognition vocabularies among college students. *Journal of Educational Psychology*, 1946, **37**, 436-439.
- Hausman, J. Research on teaching the visual arts. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand-McNally, 1963.
- Hendrickson, G. Music. In C. W. Harris (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1960.
- Henmon, V. A. C. *A French word book based on a count of 400,000 running words* (University of Wisconsin Bureau of Educational Research Bulletin, No. 3). Madison: University of Wisconsin, 1924.
- Henmon, V. A. C., & Stagner, R. Psychology of the school subjects: Foreign languages. *Review of Educational Research*, 1931, **1**, 371-378, 421-428.
- Hewitt, T. B. For briefer text-vocabularies. *German Quarterly*, 1936, **9**, 25-26.
- Hildreth, G. All in favor of a low vocabulary . . . *Elementary School Journal*, 1943, **43**, 462-470.
- Hildreth, G. A comparison of the Dale, Dolch and Rinsland word lists. *Journal of Educational Psychology*, 1948, **39**, 40-46. (a)
- Hildreth, G. Word frequency as a factor in learning to read and spell. *Journal of Educational Research*, 1948, **41**, 467-471. (b)
- Hildreth, G. *Teaching spelling: A guide to basic principles and practices*. New York: Holt, 1955.

- Hilliard, G. H. Probable types of difficulties underlying low scores in comprehension tests. *University of Iowa Studies in Education*, 1924, **2**(6).
- Hockett, J. A. The vocabularies and contents of elementary school readers. *Bulletin of Department of Education, State of California*, May 1, 1938.
- Hockett, J. A., & Neeley, D. P. A comparison of the vocabularies of thirty-three primers. *Elementary School Journal*, 1936, **37**, 190-202.
- Horn, E. A basic writing vocabulary: 10,000 words most commonly used in writing. *University of Iowa Monographs in Education*, 1926, **4**.
- Horn, E. *Methods of instruction in the social studies*. (Report of the Commission on the Social Studies, Part 15.) New York: Scribner, 1937.
- Horn, E. The validity and reliability of adult vocabulary lists. *Elementary English Review*, 1939, **16**, 129-134, 138.
- Horn, E. Spelling. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1941.
- Horn, E. Spelling. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1950.
- Horn, E., & Peterson, T. *The basic vocabulary of business letters*. New York: Gregg, 1943.
- Horn, M. D. *A study of the vocabulary of children before entering the first grade*. Baltimore: International Kindergarten Union, 1928.
- Horn, T. D. Spelling. In R. L. Ebel (Ed.), *Encyclopedia of educational research* (4th ed.). New York: Macmillan, 1969.
- Horner, V. *Music education: The background of research and opinion*. Hawthorne, Victoria: Australian Council for Educational Research, 1965.
- Horrocks, J. E., & Schoonover, T. I. *Measurement for teachers*. Columbus, Ohio: Merrill, 1968.
- Huey, E. B. *The psychology and pedagogy of reading*. New York: Macmillan, 1908.
- Humphrey, C. W. Is word frequency misused? *Business Education World*, 1945, **25**, 467-469.
- Hunnicut, C. W., & Iverson, W. J. (Eds.). *Research in the three R's*. New York: Harper, 1958.
- In honor of Edward Lee Thorndike. *Teachers College Record*, 1926, **27**, 458-586.
- Irmia, Sister M. An evaluation of the vocabulary content of twelve series of primary readers. *Catholic University of America Educational Research Bulletin*, 1929, **4**, 1-52.
- James, L. B., & Crape, L. *Geography for today's children*. New York: Appleton-Century-Crofts, 1968.
- Jenson, L. *Evaluation of reading difficulty in encyclopedia contents in Grades 5-8*. Unpublished master's thesis, University of Minnesota, 1939.
- Johns, J. L. A list of basic sight words for older disabled readers. *English Journal*, 1972, **61**, 1057-1059.
- Johns, J. L. Should the Dolch list be retired, replaced, or revised? *Elementary School Journal*, 1974, **74**, 375-380.
- Johnson, D. D. The teaching of reading. *Journal of Educational Research*, 1974, **67**, 412-420.
- Johnson, M. E. The vocabulary difficulty of content subjects in grade five. *Elementary English*, 1952, **29**, 277-280.
- Johnson, R. H., & Bond, G. L. Reading ease of commonly used tests. *Journal of Applied Psychology*, 1950, **34**, 319-324.
- Johnston, M. *Cognate relationships between English and Spanish vocabularies as a basis for instruction*. Unpublished doctoral dissertation, University of Texas, 1939.

- Joncich, G. *The sane positivist: A biography of Edward L. Thorndike*. Middletown, Conn.: Wesleyan University Press, 1968.
- Jones, A. N., & Evans, G. K. Areas of needed research in music education. *Education*, 1951, **72**, 23-27.
- Jones, W. F. *Concrete investigations of the materials of English spelling*. Vermillion: University of South Dakota, 1915.
- Joos, M. Language and the school child. *Harvard Educational Review*, 1964, **34**, 203-210.
- Judd, C. H. *Psychology of social institutions*. New York: Macmillan, 1927.
- Judd, C. H. The significance for textbook-making of the newer concepts in education. *Elementary School Journal*, 1936, **36**, 575-582.
- Kaestle, C. F. *The evolution of an urban school system: New York City, 1750-1850*. Cambridge, Mass.: Harvard University Press, 1973.
- Kavetsky, J., & Morrison, J. C. English as a second language. In C. W. Harris (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1960.
- Kearney, N. C. An analysis of the vocabularies of first grade reading material. *Journal of Educational Research*, 1950, **43**, 481-493.
- Kelley, A. A study of eight primary readers. *British Journal of Educational Psychology*, 1954, **24**, 49-50.
- Kelley, T. L. *Crossroads in the mind of man*. Stanford, Calif.: Stanford University Press, 1928.
- Kelley, T. L., & Krey, A. C. *Tests and measurements in the social sciences*. (Report of the Commission on the Social Studies, Part 4.) New York: Scribner, 1934.
- Kennon, L. H. V. Tests of literacy vocabulary for teachers of English. *Teachers College Contributions to Education*, 1926, **223**.
- Kepner, T. Vocabulary versus content in junior high school social studies. *Historical Outlook*, 1929, **20**, 30-33.
- Kirkpatrick, E. A. Number of words in an ordinary vocabulary. *Science*, 1891, **18**, 107-108.
- Kirkpatrick, E. A. A vocabulary test. *Popular Science Monthly*, 1907, **70**, 157-164.
- Kirkpatrick, J. J., & Cureton, E. E. Vocabulary item difficulty and word frequency. *Journal of Applied Psychology*, 1919, **33**, 347-351.
- Klare, C. R. *The measurement of readability*. Ames: Iowa State University Press, 1963.
- Klare, G. R., Mabry, J. E., & Gustafson, L. M. The relationship of style difficulty to immediate retention and to acceptability of technical materials. *Journal of Educational Psychology*, 1955, **46**, 287-295.
- Knott, T. A. Observations on vocabulary problems. *Elementary English Review*, 1940, **17**, 63-67.
- Kolevzon, E. R., & Maloff, R. *Vitalizing geography in the classroom*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.
- Krug, E. A. *The shaping of the American high school*. New York: Harper & Row, 1964.
- Krug, E. A. *The shaping of the American high school, 1920-1941*. Madison: University of Wisconsin Press, 1972.
- Kučera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.
- Kuhn, T. S. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.
- L. Rue, D. W. *The child's mind and the common branches*. New York: Macmillan, 1924.

- Learned, W. S. *Quality of the educational process in the United States and Europe* (Bulletin No. 20). New York: Carnegie Foundation for the Advancement of Teaching, 1927.
- Lee, J. M., & Lee, D. M. *The child and his curriculum*. New York: Appleton-Century, 1940-1960.
- Leslie, L. A. Basic shorthand vocabularies compared. *Business Education World*, 1944-1945, **25**, 115-116, 182-184, 263-266.
- Leslie, L. A. Shorthand significance of the Rinsland vocabulary study. *Business Education World*, 1945, **26**, 207-210.
- Leslie, L. A. Thorndike's 30,000 and the shorthand teacher. *Business Education World*, 1946, **26**, 526-527.
- Lessenberry, D. D., & Robinson, J. W. A basic vocabulary for typewriting. *Balance Sheet*, 1968, **49**, 388-392.
- Levi, H. S. Ineffective geography teaching—Why? *Journal of Geography*, 1938, **37**, 185-188.
- Levin, H. Reading research: What, why, and for whom? *Elementary English*, 1966, **43**, 138-147.
- Lively, B. A., & Pressey, S. L. A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision*, 1923, **9**, 389-398.
- Livengood, W. W., & Loveland, G. To whom it may concern: So you want to write a textbook? *Phi Delta Kappan*, 1952, **33**, 278-282.
- Loban, W. *The language of elementary school children*. Champaign, Ill.: National Council of Teachers of English, 1963.
- Lodge, G. *The vocabulary of high school Latin*. New York: Teachers College, 1907.
- Lorge, I. The English semantic count. *Teachers College Record*, 1937, **39**, 65-77.
- Lorge, I. Thorndike's contribution to the psychology of learning of adults. *Teachers College Record*, 1940, **41**, 778-788.
- Lorge, I. Word lists as background for communication. *Teachers College Record*, 1944, **45**, 543-552.
- Lorge, I., & Chall, J. Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *The Journal of Experimental Education*, 1963, **32**, 147-157.
- Lynch, J. J., & Eyans, B. *High school English textbooks: A critical examination*. Boston: Little, Brown, 1963.
- MacGinitie, W. M. Language development. In R. L. Ebel (Ed.), *Encyclopedia of educational research* (4th ed.). New York: Macmillan, 1969.
- MacLachy, J. H., & Wardwell, F. Common pre-primer words. *Educational Research Bulletin*, 1948, **27**, 199-206, 226.
- Mallinson, G. G. Some problems of vocabulary and reading difficulty in teaching junior high school science. *School Science and Mathematics*, 1952, **52**, 269-274.
- Mathews, C. O. The grade placement of curriculum materials in the social studies. *Teachers College Contributions to Education*, 1926, **241**.
- Mayer, M. The trouble with textbooks. *Harper's Magazine*, July 1962, pp 65-71.
- McAulay, J. D. E. *Trends in elementary school geography, 1928-1948*. Unpublished doctoral dissertation, Stanford University, 1948.
- McAulay, J. D. Whither elementary school geography. *Journal of Geography*, 1951, **50**, 189-192.
- McCall, W. A. *How to measure in education*. New York: Macmillan, 1923.
- McCall, W. A. *Measurement*. New York: Macmillan, 1939.
- McCarthy, D. Language development. In C. Murchison (Ed.), *A handbook of child psychology* (2nd ed.). Worcester, Mass.: Clark University Press, 1933.
- McKee, G. Children's themes as a source of spelling vocabulary. *Elementary School Journal*, 1924, **25**, 197-206.

- McKee, P. Word lists and vocabulary difficulty in reading matter. *Elementary English Review*, 1937, **14**, 241-245.
- Miles, J. Eliminating the guessing factor in the multiple choice test. *Educational and Psychological Measurement*, 1973, **33**, 637-651.
- Miller, H. A. *The school and the immigrant*. Cleveland: Cleveland Foundation, 1906.
- Mills, R. E., & Richardson, J. R. What do publishers mean by "grade level"? *Reading Teacher*, 1963, **16**, 359-362.
- Mitchell, D. Children's ability and educational opportunities. In J. C. Bell (Ed.), *Contributions to education* (Vol. 1). Yonkers, N.Y.: World Book, 1924.
- Monmouth, R. A., Asbell, A. E., Jenkins, H. B., & Pieters, C. E. Vocabulary difficulties. *Educational Research Bulletin*, 1924, **3**, 279-281.
- Moore, J. E. Art education. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1941.
- Morgan, C. L., & Bonham, D. N. Difficulty of vocabulary learning as affected by parts of speech. *Journal of Educational Psychology*, 1944, **35**, 369-377.
- Morphett, M. V., & Washburne, C. (Eds.). *The right book for the right child* (3rd ed.). New York: Day, 1942.
- Morris, I. *The teaching of English as a foreign language*. New York: Macmillan, 1947.
- Morrison, J. C. New York State Regents literacy test. *Journal of Educational Research*, 1925, **12**, 145-155.
- Morrison, V. M. Children can read the textbook. *Journal of Geography*, 1962, **61**, 317-320.
- Moser, K. W. *A vocabulary study of textbooks in American history for the senior high school*. Unpublished master's thesis, University of Chicago, 1927.
- Mowrer, O. H. The psychologist looks at language. *American Psychologist*, 1954, **9**, 660-694.
- Murra, W. F., Wesley, C. B., & Zink, N. E. Social studies. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1941.
- National Council of Teachers of Mathematics. *The teaching of algebra*. (Seventh Yearbook.) Washington, D.C.: Author, 1932.
- Nelson, A. H. Effects of the Depression on textbooks provided in the schools. *Elementary School Journal*, 1935, **36**, 17-25.
- Newport, J. F. The readability of science textbooks for elementary school. *Elementary School Journal*, 1965, **66**, 40-43.
- Nicholson, A. *A speller for the use of the teachers of California*. Sacramento, Calif.: State Printing Office, 1914.
- Nolte, K. F. *The effect of simplification of vocabulary on comprehension in reading*. Unpublished doctoral dissertation, State University of Iowa, 1936.
- O'Connor, J., & Filley, M. E. A junior English vocabulary test. *Personnel Journal*, 1933, **12**, 204-212.
- Ogden, C. K. *The system of basic English*. New York: Harcourt, Brace, 1934.
- Ojemann, R. The reading ability of parents and factors associated with the reading difficulty of parent education materials. In G. D. Stoddard (Ed.), *Researches in parent education, II*. Iowa City: University of Iowa, 1934.
- Otis, A. S. Considerations concerning the making of a scale for the measurement of reading ability. *Pedagogical Seminary*, 1916, **23**, 528-549.
- Otto, W. Thorndike's *Reading as reasoning*: Influence and impact. *Reading Research Quarterly*, 1971, **6**, 435-448.
- Overholser, O. V. *Grade placement analysis of reference books*. Unpublished master's thesis, Colorado State College of Education, 1935.
- Palmer, H. E. *The scientific study and teaching of languages*. New York: World Book, 1917.

- Patty, W. W., & Painter, W. I. Improving our methods of selecting high-school textbooks. *Journal of Educational Research*, 1931, **24**, 23-32, 127-134.
- Percival, W. P. *A study of the causes and subjects of school failure*. Unpublished doctoral dissertation, Teachers College, 1926.
- Phillips, M. V. Methods to improve skill in the use of a geography vocabulary at the high school level. *Journal of Geography*, 1956, **55**, 369-374.
- Piaget, J. *The language and thought of the child* (M. Warden, trans.). New York: Harcourt, Brace, 1926.
- Powers, S. R. The vocabularies of high school science textbooks. *Teachers College Record*, 1925, **26**, 368-382.
- Powers, S. R. The vocabulary of scientific terms for high school teachers. *Teachers College Record*, 1926, **28**, 220-245.
- Pressey, L. C. The determination of the technical vocabularies of the school subjects. *School and Society*, 1924, **20**, 91-96. (a)
- Pressey, L. C. *The technical vocabularies of the public school subjects*. Bloomington, Ill.: Public School Publishing Company, 1924. (b)
- Pressey, L. C. Fundamental vocabulary in elementary school geography. *Journal of Geography*, 1933, **32**, 78-81.
- Pressey, L. C. A study in the learning of the fundamental special vocabulary of history from the fourth through the twelfth grades. In T. L. Kelley & A. C. Krey (Eds.), *Tests and measurements in the social sciences*. New York: Scribner, 1934.
- Pressey, L. C., & Elam, S. K. The fundamental vocabulary of elementary school arithmetic. *Elementary School Journal*, 1932, **33**, 46-50.
- Rader, G. The vocabulary burden of a junior high school textbook in biology. *Educational Research Bulletin*, 1922, **1**, 231-232.
- Reed, H. B. *Psychology of elementary school subjects* (Rev. ed.). Boston: Ginn, 1938.
- Rejall, A. E. Reading vocabularies. *Thirty and one reading tests for voters and citizenship*. New York: Noble & Noble, 1926.
- Rice, J. M. The futility of the spelling grind. *The Forum*, 1897, **23**, 163-172.
- Richey, H. G. The professional status of textbook authors. In G. M. Whipple (Ed.), *The textbook in American education*. (Thirtieth Yearbook of the National Society for the Study of Education, Part 2.) Bloomington, Ill.: Public School Publishing Company, 1931.
- Rinsland, H. D. Analysis of completion sentences and arithmetical problems as items for intelligence tests. *Teachers College Contributions to Education*, 1935, **666**.
- Rinsland, H. D. *Constructing tests and grading in elementary and high school subjects*. New York: Prentice-Hall, 1937.
- Rinsland, H. D. *A basic vocabulary of elementary school children*. New York: Macmillan, 1945.
- Roeber, E. C. A comparison of seven interest inventories with respect to word usage. *Journal of Educational Research*, 1948, **42**, 8-17.
- Ross, D. G. *Stanley Hall: The psychologist as prophet*. Chicago: University of Chicago Press, 1972.
- Rowe, C. E. The writing of infrequently used words in shorthand. *Teachers College Contributions to Education*, 1943, **869**.
- Rudy, W. *Schools in an age of mass culture*. Englewood Cliffs, N. J.: Prentice-Hall, 1965.
- Russell, D. H. Reading research that makes a difference. *Elementary English*, 1961, **38**, 74-78.
- Russell, D. H., & Fea, H. R. Research on teaching reading. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand-McNally, 1963.

- St. Lawrence, F. Teaching aids in biology texts. *Phi Delta Kappan*, 1952, **33**, 288-289.
- Scarlet, W. Remedial reading in the New York City high schools. *High Points*, 1938, **20**, 19-24.
- Schauer, V. P. Teaching children to read the geography textbook. *Journal of Geography*, 1945, **44**, 279-287.
- Schorling, R., & Edmonson, J. B. The technique of textbook authors. In G. M. Whipple (ed.), *The textbook in American education*. (Thirtieth Yearbook of the National Society for the Study of Education, Part 2). Bloomington, Ill.: Public School Publishing Company, 1931.
- Schramm, W. The publishing process. In L. J. Cronbach (Ed.), *Text materials in modern education*. Urbana: University of Illinois Press, 1955.
- Schwesinger, G. C. The social-ethical significance of vocabulary. *Teachers College Contributions to Education*, 1921, **221**. New York: Teachers College, 1926.
- Scott, F., & Myers, G. C. Children's empty and erroneous conceptions of the commonplace. *Journal of Educational Research*, 1923, **8**, 327-334.
- Seago, D. W. An analysis of language factors in intelligence tests. *Mental Measurement Monographs*, 1925, **1**, 1-125.
- Seashore, R. H. The measurement and analysis of extent of vocabulary. *Psychological Bulletin*, 1933, **30**, 709-710.
- Seashore, R. H., & Eckerson, L. D. The measurement of individual differences in general English vocabulary. *Journal of Educational Psychology*, 1940, **31**, 14-38.
- Seegers, J. C. Vocabulary problems in the elementary school: A digest of current research. *The Elementary English Review*, 1939, **16**, 157-166, 199-204, 234-239, 279-282, 320-326.
- Seegers, J. C., & Seashore, R. H. How large are children's vocabularies? A discussion. *Elementary English*, 1949, **26**, 181-194.
- Serra, M. C. Amplifying and simplifying instructional materials: Effect on comprehension. *Elementary School Journal*, 1954, **55**, 77-81.
- Shapiro, S. R. A remedial project in the social studies. *High Points*, 1938, **20**, 9-13.
- Sherman, L. A. *Analytics of literature*. Boston: Ginn, 1893.
- Sholty, M. A study of the reading vocabulary of children. *Elementary School Teacher*, 1912, **12**, 272-277.
- Showel, M. A comparison of alternative media for teaching beginning typists. *Journal of Educational Research*, 1974, **67**, 279-285.
- Skinner, M. Our low IQ's—What shall we do with them? In J. C. Bell (Ed.), *Contributions to education* (Vol. 1). Yonkers, N. Y.: World Book, 1924.
- Smith, K. J., & Heddens, J. W. The readability of experimental mathematics materials. *Arithmetic Teacher*, 1964, **11**, 391-394.
- Smith, M. E. An investigation of the development of the sentence and the extent of vocabulary in young children. *University of Iowa Studies in Child Welfare*, 1926, **3**(5), 1-93.
- Smith, M. K. Measurement of size of general English vocabulary through elementary grades and high school. *Genetic Psychological Monographs*, 1941, **24**, 311-345.
- Smith, N. B. *American reading instruction*. Newark, Del.: International Reading Association, 1965.
- Social Science Citation Index*. Philadelphia: Institute for Scientific Information, 1974.
- Stauffer, R. G. Thorndike's *Reading as reasoning*: A perspective. *Reading Research Quarterly*, 1971, **6**, 443-448.
- Stern, C., & Stern, W. L. *Die Kindersprache* [Children's Speech]. Leipzig: Barth, 1907.
- Stevenson, P. R. Difficulties in problem solving. *Journal of Educational Research*, 1925, **11**, 95-103.

- Strang, R., Checkovitz, A., Gilbert, C., & Scoggin, M. *Gateways to readable books*. New York: Wilson, 1944.
- Strickland, C. E., & Burgess, C. (Eds.). *Health, growth, and heredity: G. Stanley Hall on natural education*. New York: Teachers College, 1965.
- Strickland, R. G. The language of elementary school children: Its relationship to the language of reading textbooks and the quality of reading of selected children. *Bulletin of the School of Education, Indiana University*, 1962, **38**, 1-131.
- Taylor, W. L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, **30**, 415-433.
- Temp, G. Literature in the secondary school. *Review of Educational Research*, 1964, **34**, 195-202.
- Terman, L. M. *The measurement of intelligence*. Boston: Houghton Mifflin, 1916.
- Terman, L. M. The vocabulary test as a measure of intelligence. *Journal of Educational Psychology*, 1918, **9**, 452-466.
- Terman, L. M., & Childs, H. G. A tentative revision and extension of the Binet-Simon measuring scale of intelligence. *Journal of Educational Psychology*, 1912, **3**, 198-208.
- Tharp, J. B. Michael West of Dacca, India. *Educational Research Bulletin*, 1931, **10**, 34-36.
- Tharp, J. B., Coleman, A., & King, C. B. Foreign languages, modern. in W. S. Monroe (Ed.), *Encyclopedia of educational research* (Rev. ed.). New York: Macmillan, 1950.
- Thorndike, E. L. The elimination of pupils from schools. *United States Bureau of Education Bulletin*, 1907, **4** (Whole No. 379).
- Thorndike, E. L. The measurement of ability in reading: Preliminary scales and tests. *Teachers College Record*, 1914, **15**, 207-277.
- Thorndike, E. L. The measurement of achievement in reading: Word knowledge. *Teachers College Record*, 1916, **17**, 430-451.
- Thorndike, E. L. Word knowledge in the elementary school. *Teachers College Record*, 1921, **22**, 334-370.
- Thorndike, E. L. *The psychology of algebra*. New York: Macmillan, 1923.
- Thorndike, E. L. *The teacher's word book*. New York: Teacher's College, 1927. (Originally published, 1921.)
- Thorndike, E. L. *Adult learning*. New York: Macmillan, 1928.
- Thorndike, E. L. *A teacher's word book of the twenty thousand words found most frequently and widely in general reading for children and young people*. New York: Teachers College, 1932. (Originally published, 1931.)
- Thorndike, E. L. Improving the ability to read. *Teachers College Record*, 1934, **36**, 1-19.
- Thorndike, E. L. *Adult interests*. New York: Macmillan, 1935.
- Thorndike, E. L. On the number of words of any given frequency of use. *Psychological Record*, 1937, **1**, 399-406. (a)
- Thorndike, E. L. The vocabulary of books for children in Grades 3 to 8. *Teachers College Record*, 1937, **38**, 196-205, 316-323, 416-429. (b)
- Thorndike, E. L. The teaching of English suffixes. *Teachers College Contributions to Education*, 1941, **847**.
- Thorndike, E. L. *The Thorndike-Century junior dictionary*. Chicago: Scott, Foresman, 1942. (Originally published, 1935.)
- Thorndike, E. L. On the frequency of semantic changes in modern English. *Journal of General Psychology*, 1948, **39**, 23-27.
- Thorndike, E. L., & Lorge, I. The vocabulary of the *Encyclopaedia Britannica*. *Teachers College Record*, 1937, **38**, 485-490.

- Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College Press, 1972. (Originally published, 1944.)
- Thorndike, E. L., & Symonds, P. Difficulty, reliability, and grade achievements in a test of English words. *Teachers College Record*, 1925, **24**, 438-445.
- Thorndike, E. L., & Wohlfarth, J. H. *Growth in spelling* (2 vols.). Yonkers, N. Y.: World Book, 1929.
- Thorndike, R. L., & Hagen, E. *Measurement and evaluation in psychology and education* (2nd ed.). New York: Wiley, 1961.
- Thorpe, L. P. Learning theory and music teaching. In N. B. Henry (Ed.), *Basic concepts in music education*. (Fifty-seventh Yearbook of the National Society for the Study of Education, Part I.) Chicago: University of Chicago Press, 1958.
- Thralls, Z. A. *The teaching of geography*. New York: Appleton-Century-Crofts, 1958.
- Tireman, L. S. Bilingual children. *Review of Educational Research*, 1941, **11**, 340-352.
- Tireman, L. S., Dixon, N., & Cornelius, V. Vocabulary acquisition of Spanish-speaking children. *Elementary English Review*, 1935, **12**, 118-119, 144.
- Traxler, A. E. Remedial reading in the New York City high schools. *High Points*, 1938, **20**, 7-18.
- Troen, S. K. *The public and the schools: Shaping the St. Louis system, 1838-1920*. Columbia: University of Missouri Press, 1975.
- Tuinman, J. J. Thorndike revisited—Some facts. *Reading Research Quarterly*, 1971, **7**, 195-202.
- Twaddell, W. F. Linguists and language teachers. In C. J. Kreidler (Ed.), *On teaching English to speakers of other languages*. Champaign, Ill.: National Council of Teachers of English, 1966.
- Tyack, D. B. *Turning points in American educational history*. Waltham, Mass.: Blaisdell, 1967.
- Tyack, D. B. *The one best system: A history of American urban education*. Cambridge, Mass.: Harvard University Press, 1974.
- Uhl, W. L. Scientific determination of the content of the elementary course in reading. *University of Wisconsin Studies in the Social Sciences and History*, 1921, **4**.
- Uhl, W. L. *The materials of reading: Their selection and organization*. New York: Silver, Burdett, 1924.
- Uhl, W. L. The materials of reading. In G. M. Whipple (Ed.), *The teaching of reading: A second report*. (Thirty-sixth Yearbook of the National Study for the Study of Education, Part I.) Bloomington, Ill.: Public School Publishing Company, 1937.
- Upton, L. How to use the textbook. *Journal of Geography*, 1933, **32**, 359-364.
- Vander Beke, C. E. *French word book*. New York: Macmillan, 1929.
- Vogel, M., & Washburne, C. An objective method of determining grade placement of children's reading material. *Elementary School Journal*, 1928, **28**, 373-381.
- Wang, C. K. A. *An annotated bibliography of mental tests and scales*. Peiping, China: Catholic University Press, 1939.
- Ward, C. F. *Minimum French vocabulary test book*. New York: Macmillan, 1926.
- Ward, J. L., & Stevenson, P. R. The vocabulary of a history text. *The American School Board Journal*, 1925, **71**, 65, 131.
- Ware, A. E. Review of *California* in "Progressive geographies series." *Journal of Geography*, 1921, **20**, 118.
- Washburne, C. W. A spelling curriculum based on research. *Elementary School Journal*, 1923, **23**, 751-762.

- Washburne, C. W. An autobiographical sketch. In R. J. Havighurst (Ed.), *Leaders in American education*. (Seventieth Yearbook of the National Society for the Study of Education, Part 2.) Chicago: University of Chicago Press, 1971.
- Washburne, C., & Morphet, M. V. Grade placement of children's books. *Elementary School Journal*, 1938, **38**, 355-364.
- Watters, G. Local word book reduces need for teacher help in early grades. *Nation's Schools*, 1936, **17**, 37-39.
- West, L. J. Business education. In R. L. Ebel (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1969.
- West, L. J. Research on teaching business subjects. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand-McNally, 1973.
- Whipple, G. Procedures used in selecting schoolbooks. *Elementary School Journal*, 1936, **36**, 665-673, 760-775; **37**, 47-57.
- Whipple, G. M. (Ed.). *Report of the National Committee on Reading*. (Twenty-fourth Yearbook of the National Society for the Study of Education, Part 1.) Bloomington, Ill.: Public School Publishing Company, 1925.
- Whipple, G. M. Needed investigations in the field of the textbook. *Elementary School Journal*, 1935, **35**, 575-582.
- Whitbeck, R. H. Thirty years of geography in the United States. *Journal of Geography*, 1921, **20**, 121-128.
- Wiley, W. Difficult words and the beginner. *Journal of Educational Research*, 1928, **17**, 278-289.
- Wilson, E. M. Review of the Barrows and Parker geography series. *Journal of Geography*, 1937, **36**, 331.
- Wirth, A. G. *Education in the technical society*. Scranton, Pa.: Intext, 1972.
- Wise, C. T. Selection and gradation of words in spelling. *Elementary School Journal*, 1934, **34**, 754-766.
- Wise, C. T. Selection and gradation of words in spelling. *Elementary School Journal*, 1935, **36**, 281-289.
- Wishy, B. *The child and the republic: The dawn of modern American child nurture*. Philadelphia: University of Pennsylvania Press, 1968.
- Wrightstone, J. W., Justman, J., & Robbins, I. *Evaluation in modern education*. New York: American Book Company, 1956.
- Yoakam, G. A. A cooperative experiment in word pronunciation. *University of Pittsburgh School of Education Journal*, 1927, **2**, 49-55.
- Yoakam, G. A. The reading difficulty of school textbooks. *Elementary English Review*, 1945, **22**, 304-309, 333-336.
- Yoakam, G. A. (Ed.). *Reading and the language arts in the school program*. (A Report of the Sixth Annual Conference on Reading, University of Pittsburgh.) Pittsburgh: University of Pittsburgh, 1950.
- Yoakam, G. A. Unsolved problems in reading. *Elementary English*, 1954, **31**, 427-430.
- Young, J. W. The work of the national committee on mathematical requirements. *Mathematics Teacher*, 1921, **14**, 5-15.
- Zink, N. E. Geography aids from the research field. *Educational Method*, 1938, **17**, 301-305.

Chapter 3

THE CONTRIBUTIONS OF B. F. SKINNER TO EDUCATION AND SOME COUNTERINFLUENCES

by

ROBERT GLASER

University of Pittsburgh

1. THE RELATIONSHIP OF PSYCHOLOGY TO EDUCATION

ANY profession relating to the human condition rests its practices in some manner on a foundation of its beliefs about human nature. At one time, economists developed economic theory on the basis of a metaphorical model of the rational man; psychoanalysts rested their practices on the psychoanalytic description of human development. Educators in America, especially over the past 75 years, have searched for their particular set of principles of human nature and have found it in various forms. William James, G. Stanley Hall, Edward Thorndike, John Dewey, B. F. Skinner, and Jean Piaget all provided, in one way or another, principles of human behavior from which educational practice could be interpreted. At the least, these individuals pointed out that the scientific psychological study of humans was fundamental to establishing a disciplinary base for the development of educational practice. A philosophy of education was a necessary but insufficient condition. Knowledge of human behavior verified through the methods of science was also required. Scientific knowledge and convictions about human values and social objectives combine to yield an effective educational profession.

In American education, activities centering around the complex of interrelationships involved in translations of scientific knowledge into practice and of practice into scientific knowledge have had an erratic history. At the beginning of this century those great figures concerned with the development of scientific psychology imported from the European laboratories were enthusiastically encouraging active intercourse between science and its application to education. And the kinds of theory

that appealed to them, as compared to others, were, to some extent, a function of a theory's practical serviceability. Of the psychologies that were available, the one that became most popular was related to the British associationism, which appeared to be the most pragmatic and was seen as being readily put to the service of education. In a book entitled *Association Theory To-day*, Robinson (1932) makes this point:

British thought, until well on into the nineteenth century, was strongly associational, but the associationism of our times is not to be understood as simply a continuation of the British tradition. The emphasis that our writers have put upon association requires for its explanation reasons why the conception has had a peculiar fitness for the American environment. Perhaps because the American public has had an eager faith in the practical advantages of learning, the American scholar has felt a great cail to be useful. Our philosophy has tried several doctrines, but pragmatism has been its one whole-hearted theory. Our physics has been a little mathematical, but very experimental. Our psychology has had its doctrines of sensation and of instinct, but these have been minority matters as compared with association, which opens up the school room. We can see why, then, among the numerous conceptions that might have dominated psychological thought, association should have had a genuine appeal. The texture of experience is something about which nothing can be done. Innate capacities are merely the limits within which an energetic improver of humanity can work. But association strikes at the heart of education. It holds a promise that changes can be worked in human nature, because it is, in fact, a theory of such changes. (pp. 4-5)

Edward Thorndike's experiments and theory of learning, and his attempts at application to education, fit into this pattern. In the preface of his influential book entitled *The Psychology of Arithmetic* (1922), he wrote as follows:

Within recent years there have been three lines of advance in psychology which are of notable significance for teaching. The first is the new point of view concerning the general process of learning. We now understand that learning is essentially the formation of connections or bonds between situations and responses, that the satisfyingness of the result is the chief force that forms them, and that habit rules in the realm of thought as truly and as fully as in the realm of action.

The second is the great increase in knowledge of the amount, rate, and conditions of improvement in those organized groups or hierarchies of habits which we call abilities, such as ability to add or ability to read. Practice and improvement are no longer vague generalities, but concern changes which are definable and measurable by standard tests and scales.

The third is the better understanding of the so-called "higher processes" of analysis, abstraction, the formation of general notions, and reasoning. The older view of a mental chemistry whereby sensations were compounded into percepts, percepts were duplicated by images, percepts and images were amalgamated into abstractions and concepts, and these were manipulated by reasoning, has given way to the understanding of the laws of response to elements or aspects of situations.

This book presents the applications of this newer dynamic psychology to the teaching of arithmetic. (pp. v-vi)

In this book, Thorndike applied his theory and findings about learning directly to the teaching process. The theory of stimulus-response bonds that made up complex chains of behavior was applied to the analysis of arithmetic tasks; the task of adding integers, for example, was carefully analyzed in terms of S-R bonds that could be taught and observed by the teacher. Thorndike also applied the results of his experimental work on transfer of training and reward in suggestions for practical teaching techniques.

1.1 *Contrasting Approaches*

John Dewey is a contrast to Thorndike not only with respect to his beliefs about the nature of human behavior but also with respect to the relationship between psychology and its application. Dewey (1900) said the following in his presidential address before the American Psychological Association in 1899:

"Do we not lay a special linking science everywhere else between the theory and practical work? We have engineering between physics and the practical workingmen in the mills; we have a scientific medicine between the natural science and the physician" (p. 138).¹ The sentences suggest . . . that the real essence of the problem is found in . . . [a] connection between the two extreme terms—between the theorist and the practical worker—through the medium of the linking science. The decisive matter is the extent to which the ideas of the theorist actually project themselves, through the kind offices of the middleman, into the consciousness of the practitioner. It is the participation by the practical man in the theory, through the agency of the linking science, that determines at once the effectiveness of the work done, and the moral freedom and personal development of the one engaged in it. (pp. 110–111)

While the distinction can be belabored, it seems that a difference between Dewey and Thorndike emerges on the subject of the translation of science into practice. Dewey envisioned a special structure, a linking science, that intervened between scientific theory and practical application to give (to those servicing the practitioner) some conceptual framework into which knowledge could be cumulated and corrected. Thorndike, on the other hand, operated in a more direct fashion. There were certain laws of learning—the laws of S-R bonds, the law of effect, and findings about the specificity of transfer—that could be immediately applied to the design of instruction and materials for teaching.

Dewey apparently elected to leave the translation of science into practice to the long-term venture that he conceived it to be. While being a firm advocate of psychological research, he thought it most immediately important to move from concepts to implementation, which appears to have led to his establishment of the laboratory school. In the laboratory school it was possible to provide examples of practice and demonstrate their

¹ Dewey is quoting Hugo Munsterberg (1899).

feasibility so that others could use them as a basis for further implementation of new ideas. The underlying principles of human behavior involved were only very generally spelled out. In Dewey's thinking, the findings of science would eventually find their way into practical application, but the immediate development of schools needed to proceed on the basis of more intuitive principles of learning. The fact that these intuitive principles should provide important concepts for laboratory study and for the development of theories of learning and child development was not particularly entertained. This approach has the disadvantage of generating little force either to move scientific knowledge into practice or to have the principles of schooling interpreted as scientific concepts for study; an interrelating theory could hardly emerge under these conditions.

Thorndike, on the other hand, came armed with tested principles from the laboratory and from theory that could be applied directly to the design of learning conditions and teaching practices in the classroom. Thorndike's approach set a particular pattern: the combination in one person of the theoretical scientist and the applied scientist interested in designing conditions for effective teaching and learning. Direct application of this kind appears to require someone trained in the science of psychology who is motivated to look at problems in education. It is a highly individualistic enterprise that, to some extent, does get things done; the ideas involved are built into the design of educational materials and practices. For example, Thorndike's notions led to some textbook redesign in mathematics, to new educational psychology texts, and to some general curriculum changes. But these specific events do not necessarily, and did not, contribute to the development of the generally available linking science envisioned by Dewey.

1.2 *A Loosening of Ties*

The spirit of close affinity between psychology and education persisted for a short time after the time of Dewey and Thorndike; but following this period, education and psychology went their separate ways. The developing insularity of education, on the one hand, and psychology, on the other, was fostered by activities on both sides.

Education and psychology addressed what they considered to be their immediately demanding audiences and not their familial relations. Psychology aspired to become a natural science and take its place among the hard sciences, and psychologists went into the laboratory to work out experimental techniques with tasks designed more for theoretical purposes than for relevance to educational subject matter. Education, on the other hand, particularly educational psychology, found its challenges with practitioners—with the teachers and with the individuals who managed schools and designed curricula—and supported the educational professional establishment by working with problems of teaching practice in

colleges of education. This target group was reached through the growing number of colleges of education and teachers colleges throughout the country.

As psychology and education, in the main, parted company, psychologists joined the faculty of arts and sciences, and most educational psychologists joined separate departments in faculties of education. As a result, most of the individuals interested in the development of educational psychology were trained in colleges of education and were concerned primarily with their constituency of students of education. The two enterprises, the main body of experimental psychology and educational psychology, took on different characteristics.

A signal event in this climate was Pressey's (1933) educational psychology text entitled *Psychology and the New Education*. For years to come, this important book set the tone and pattern of educational psychology textbooks written for the growing colleges of education. The style of the earlier books like Thorndike's *Educational Psychology* (1923), in which laboratory experiments and theory were presented to educational psychologists, had changed. The first chapter of Pressey's preface reads as follows:

This volume might be called an experiment in subject matter. It is an attempt at a sweeping reorganization of the usual treatment of educational psychology. The aim has been not so much at a systematic presentation of psychological data and theory as at maximal usefulness in contributing to understanding of educational problems. (p. xxvii)

Pressey goes on to say:

Not only is there gross reorganization of the material in order to make the treatment more practical and functional for teacher training, but, throughout, this practical and functional point of view markedly affects the selection of material. The discussion of learning begins not with the escape of cats from a box, but with a "field study" of children cooperating in the solving of a puzzle. The learning curves show progress in school subjects. And the forgetting curves show forgetting not of nonsense syllables, but of arithmetic, history, chemistry. The writer does not question the value of animal experimentation or work in the psychological laboratory. But he feels that educational psychology should deal with children, and the learning of school materials, in school; he feels also that the subject has now progressed far enough that it can develop its topics in terms of its own subject matter.

With such changes in material and organization, the inclusion of psychological matter of a controversial character seemed undesirable. Thus the concept of instinct is eliminated without debate. The treatment of learning may seem suggestive of a gestalt point of view, while in other places there is what may seem a naive and uncritical use of the concept involved in the "law of effect." The instructor may bring in a greater refinement in the discussion of such topics, if he wishes. But the purpose of the text is pragmatic: to give a clear and straightforward interpretation of the facts under discussion in such fashion as to be of the greatest possible service to a teacher. Doubtless there has resulted much over-simplification of treatment. But the writer would prefer

such a result to a treatment so balanced or elaborate that it left a teacher in confusion on issues where she must, for practical purposes, adopt some point of view. (pp. xxvii-xxix)

The new textbooks emphasized principles that could be given to practitioners so that they had some general body of knowledge that might influence their teaching practices, and general instructional principles were presented with appropriate classroom experiments and case studies.

The direct involvement of the experimental psychologist in the development of teaching practices and instructional materials as part of an experimental applied psychology, however, was of less concern. Back at the faculty of arts and sciences, education was hardly an activity in the "true" psychology departments. The separation between psychology and education is documented in one way by noting that the 1950 yearbook of the National Society for the Study of Education entitled *Learning and Instruction* (Henry, 1950) did not list Hull, Skinner, Spence, or Tolman in its index. (The yearbook on the same topic in 1964 [Hilgard, 1964] lists them all in abundance, and the yearbook itself contains many chapters written by experimental psychologists.)

It was into this scene that Skinner appeared with his 1953 book, *Science and Human Behavior*, with a discussion of education and in 1954 at a conference on current trends in psychology with a paper entitled "The Science of Learning and the Art of Teaching," which was quickly followed by an article printed in *Science* in 1958 entitled "Teaching Machines." Also at this time, individuals who had worked with Skinner, like Bijou (1955), Ferster (1961), and Azrin and Lindsley (1956), were working with behavior disorders in children and adults in schools, clinics, and homes (see Krasner & Ullmann, 1965).

2. BACKGROUND

2.1 Skinner's System of Behavior

In 1938, Skinner published the book that laid out the fundamentals of his "system of behavior." This book, entitled *The Behavior of Organisms: An Experimental Analysis*, set forth a formulation of behavior, with accompanying experiments, with which Skinner's later work has been remarkably consistent. The wide ranging of Skinner's analyses into many levels of animal, human, individual, and social behavior remains true to the formulations described in this work. The success of his analytical procedures in leading to highly replicable experiments and the facility with which the techniques involved could be used appealed to a generation of psychologists and still continue to do so. The behavior that Skinner investigated was not unlike that previously investigated by Pavlov and Thorndike, but there were significant differences as a result of certain departures from traditional stimulus-response psychology and differences

in the procedures that were applied to the analysis of animal behavior.

Like others before him, Skinner professed to observe behavior in a naturalistic way by giving a detailed descriptive account of a sample of behavior as it unfolds in a particular situation, but he went beyond mere observation to find and analyze the variables of which behavior is a function. Skinner concentrated on discovering such variables in the cases he studied in a rigorously behavioristic way without resorting to hypothetical mental or neurological explanations. The main variable that enters into this scientific description of behavior is the external environment that acts upon the organism; modification in a part of the environment (a stimulus) influences behavior (a response).

The properties of the S-R "reflex" had been studied by Pavlov, Sherrington, and others in earlier work on reflex physiology and were carried over into Watson's stimulus-response psychology. Watson's formulation of behavior depended upon a description of a response and the stimulus that elicited it. Skinner found the description of behavior in terms of the specific relationships between eliciting stimuli and responses too restrictive and therefore inappropriate for describing major classes of behavior. A large body of behavior did not seem to be elicited "in the sense in which a cinder in the eye elicits closure of the lid." Much of the activity of the organism appeared to be original, "spontaneous" activity, and to assert that there must be an eliciting stimulus to be found was an "unsatisfactory appeal to ignorance" and constrained a science of behavior with respect to a major part of an organism's activity (Skinner, 1938, p. 19).

Consequently, an adequate behavioristic science of behavior must deal with events that occur without any observable antecedent events. This does not mean that these antecedent events do not occur, but that we are not always in a position to detect or observe them. This kind of behavior can be said to be "emitted" by the organism in contrast to behavior that is elicited by specific identifiable stimuli. "The attempt to force behavior into the simple stimulus-response formula has delayed the adequate treatment of that large part of behavior which cannot be shown to be under the control of eliciting stimuli" (Skinner, 1938, p. 20).

Skinner's system makes this significant departure from traditional stimulus-response psychology. The kind of behavior that is correlated with specific eliciting stimuli is called *respondent* behavior, that is, respondent to an eliciting stimuli. Behavior not under this kind of control is called *operant* behavior. Operant behavior is emitted by the organism without any specific eliciting stimulus. Being different from respondent behavior, it must be studied in different ways and through different behavior indices. What can be observed is the appearance of operant behavior that occurs over time at some more or less orderly rate. Changes in frequency of occurrence and rate of performance—increases and decreases in the frequency of what is learned or "unlearned"—are the

dependent variables and become the major assessments of behavior in Skinner's system.

The principal systematic independent variable of the system is "reinforcement." The general observation is that experimentally imposed or natural environmental consequences of behavior can increase or decrease the probability with which behavior occurs; the particular properties of the behavior that are acquired depend upon the details of these environmental consequences. A reinforcement can be any consequence of a response that leads to an increase in its frequency or the probability of its recurrence in similar situations in the future.

Reinforcers are defined simply in terms of their effects upon behavior. There need be no logical or permanent connection between a response and a reinforcer, and "seeing a connection" is not essential in operant conditioning. While there have been various theoretical interpretations of the mechanisms of reinforcement in terms of physiological mechanisms and various need satisfaction and incentive effects, Skinner actively avoids such reductionist explanations; while reductionist explanations may eventually be helpful, behavior and its interaction with the environment is a datum in its own right. The operational definition of reinforcement is emphasized; behavior is acquired and its occurrence regulated as a result of a contingent relationship between the response of the organism and a consequent event. Skinner's work deemphasizes the making of inferences about underlying states or processes from performance. He carries out a program that strongly emphasizes the discovery and analysis of empirical relationships between observable behaviors and controlling variables in the environment.

2.2 *Reinforcement Schedules*

It has been shown that the schedule on which reinforcement occurs strongly influences behavior and is often much more important than the nature and quantity of the reinforcer. The reinforcement of operant behavior need not take place at regular and uniform intervals; indeed, it does not occur so in ordinary life conditions. All of our behavior is not reinforced all of the time; most often, reinforcing events follow behavior on an intermittent basis. A hunter does not bring back a deer after every hunt and a writer does not sell an article every time he writes, but they continue to hunt and write. "The reinforcements characteristic of industry and education are almost always intermittent because it is not feasible to control behavior by reinforcing every response" (Skinner, 1953, p. 99). When learning occurs under two different schedules, the different effects are quite apparent. When animals are reinforced on a "ratio" schedule, where a reinforcement occurs after a certain number of responses have occurred, like every 50th response, then the animal works at a great rate and seems to be "highly motivated, persistent, and industrious." However,

if another schedule is in effect, an "interval" schedule where a schedule of reinforcement is determined by a time interval, say every 5 minutes, then immediately after reinforcement the animal will show a very low rate of responding or a pause and then there is a gradual acceleration to a high rate of response just near the time when a reinforcement becomes available. Skinner likens ratio and interval schedules to human situations; a ratio schedule is like piecework pay in industry and generates a high ratio of work; the pause (the inability to expend effort) in an interval schedule is analogous to the difficulty a student finds in working on a new assignment after there has just been a burst of speed in finishing a term paper. The influence of many different kinds of complex schedules of reinforcement has been intensively studied (Ferster & Skinner, 1957; Morse, 1966). The relationships of such reinforcing conditions to behavior labeled as "persistence," "exploration," and "inattention" have closely connected the concepts of reinforcement and motivation.

2.3 *Kinds of Reinforcers*

A reinforcer is defined as an event, a stimulus, or a state of affair; that changes subsequent behavior when it follows the behavior in time. An event is identified as a "positive reinforcer" when its presentation following or contingent upon the occurrence of a response increases the probability of the occurrence of that class of response. Responses are also strengthened by "negative reinforcers"; these consist of noxious or aversive events that are removed if the response occurs. When a response is no longer followed by a reinforcing event, it undergoes extinction. As successive responses fail to produce reinforcement, the recurrence of the response is less and less likely. Another class of contingencies is called "punishment." Punishment is defined either as the presentation of an aversive stimulus following a response or the removal of a positive reinforcer contingent upon a response.

There is a distinction made between negative reinforcers and punishers. Negative reinforcers precede a response and their occurrence is terminated by the response, terminating the unpleasant condition. In contrast, punishers are not in effect prior to the response but occur following a response. On the basis of his early experiments, Skinner (1938) reports that the effect of punishment was to temporarily suppress behavior, but that when punishment was removed, the response reappeared with a rate equal to that where no punishment had taken place. Skinner likens the effects of punishment to Freud's discovery of the surviving activity of repressed wishes; temporary inhibition does not permanently reduce a tendency to respond (1953, p. 184). Nonetheless, in the light of further research, the interpretation of the effects of punishment is complicated. Recent studies show a variety of effects depending upon the nature of the response involved and the intensity of the

punishment. In particular, the extent and duration of the reduction and suppression of behavior are a function of the intensity of the punishment and the schedule on which it takes place (see Azrin & Holz, 1966).

2.4 *Shaping and Differentiation of Responses*

Behavior is learned through a sequence of reinforcing situations. In Skinner's (1953) words:

Operant conditioning shapes behavior as a sculptor shapes a lump of clay. Although at the point the sculptor seems to have produced an entirely novel object, we can always follow the process back to the original undifferentiated lump, and we can make the successive stages by which we return to this condition as small as we wish. At no point does anything emerge which is very different from what preceded it. The final product seems to have a special unity or integrity of design, but we cannot find a point at which this suddenly appears. In the same sense, an operant is not something which appears full grown in the behavior of the organism. It is the result of a continuous shaping process. (p. 91)

By reinforcing a series of "successive approximations," a rare response can be brought to a high probability even though the organism originally had no tendency to make the given response. In a simple trial-and-error procedure, it is likely that the response would never be reinforced and thus never strengthened. The basic procedure is to observe the behavior of an organism in a given situation and determine the class of responses that the organism emits which resemble in some way the one to be trained. Closer and closer approximations to the desired behavior is then reinforced. For example, in training a pigeon to peck a particular spot, the shaping procedure is carried out as follows. When the bird turns slightly in the direction of the spot from any part of the cage, it is given food. This increases the frequency of this behavior. Reinforcement is then provided whenever it moves in the direction of the spot, reinforcing positions successively closer to the spot. Reinforcement is also provided when the pigeon moves its head slightly forward and makes pecking movements and then, finally, when the beak actually makes contact with the spot, that is, when it actually pecks the key that delivers food. In this way, a hungry bird can be trained to respond in a very short time. The original probability of the response was very low and a complex repertoire is produced that would never have appeared otherwise.

Through a series of successive approximations, initial responses that do not look like the final repertoire at all can be trained into a complex repertoire with a high probability of occurrence. Through the same process—the shaping of responses through differential reinforcement contingencies—humans learn the many complex repertoires characteristic of children and adults. Shaping occurs in various and subtle ways. The crude vocal responses of infants eventually develop into forms of behavior

that correspond more and more to the standards of adult talk. "When we walk on a deck of a ship at sea, a special set of contingencies prevails in maintaining our orientation in the gravitational field. The differential reinforcement sets up 'sea legs'. At the end of the voyage the old contingencies work a reverse change" (Skinner, 1953, p. 96). Differential reinforcement is used by individuals to shape and modify the behavior of others. The effects may be intentional or unintentional, such as when parents do not reinforce the baby talk of their children as they grow older or when a parent reinforces a child for whining and crying for attention.

2.5 *Discrimination*

Operant behavior acquires important connections with features of the surrounding environment. In the pigeon example described above, the spot-pecking response can be reinforced only when a signal light is on and not when the signal light is off (allowing the response to be extinguished when the light is off). Eventually, pecking the spot will only occur when the light is on. This state of affairs is described by a three-term contingency which produces discriminative operant behavior: The stimulus (the light) is the occasion upon which a response (pecking a spot) is followed by a reinforcement (the delivery of food). The effect upon the animal is that eventually the response is more likely to occur when the light is on. The process by which this comes about is called discrimination learning, and the probability of a response can be controlled by presenting or removing the discriminative stimulus. As an organism behaves in the world, operant behavior comes under this kind of stimulus control.

In general, discrimination learning is a process by which stimuli come to acquire selective control over behavior; particular situations set the occasion for the occurrence of behavior in that situation. Introduction or removal of the discriminative stimulus can result in control over the behavior. In a manner analogous to the refinement of responses in successive approximation, behavior can be brought under the control of increasingly detailed stimulus situations. Through such a set of conditions, an artist learns to see nuances in color never thought of by the nonartist, or a motorist learns a very broad category of similar stimulus events, such as that all colors of red signify danger. In the practical control of behavior, the introduction and removal of discriminative stimuli become an important operation that influences behavior.

Once a response is reliably evoked in the presence of a particular stimulus, control can be transferred to new stimulus conditions. A second stimulus can be brought to control the response by presenting the two stimuli together and by supplying the appropriate reinforcement in the presence of both the old and new stimuli. Over a series of trials where two stimuli are present, the stimulus in the presence of which the response is

evoked can be gradually eliminated or "faded", so that the learner is forced to depend on the remaining stimulus cue; as a result, a control of the response is transferred or extended to the new stimulus situation. In general, the initially effective stimulus can serve as a prompt that sets the occasion for appropriate behavior in the presence of to-be-learned stimuli, and new learning is not considered complete until the prompting stimulus is successfully removed and the behavior occurs reliably in the presence of the new stimulus situation.

A discrimination is learned when the organism responds differentially in different stimulus situations and does so reliably. In simple discrimination problems, this learning comes about because the learner is reinforced in the presence of one kind of stimulus event (the positive stimulus) and is not reinforced in the presence of another stimulus (the negative stimulus). Although there is initially some generalization between the two kinds of stimuli, behavior is eventually extinguished in the presence of the negative stimulus. This extinction process, it has been observed, brings about repeated frustration and inhibition as a result of nonreinforcement; as a result, the negative stimulus acquires the tendency to elicit aversive or "emotional" behavior from the subject that is built up as a result of extinction to the negative stimulus.

In order to counteract the effects of this extinction process and to alleviate the frustration that comes about as a result of making errors, a procedure has been devised in which a pigeon is taught a perfect discrimination in such a manner that it never responds to the negative stimulus—it never makes an "error" (Terrace, 1963). The express principle involved is that an optimal arrangement of a program sequence can be devised in which the learner makes few or no errors in the course of learning, and, as a result, no emotional behavior is built up to other stimuli in the situation as a result of nonreinforcement and extinction. This procedure, which incorporated the technique of "fading," was quickly extended to the teaching of discriminations in young children and retardates, such as discriminating differences in letters, size, and shape (Moore & Goldiamond, 1964; Sidman & Stoddard, 1967).

2.6 *Generalization and Abstraction*

When a response has been brought under the control of a given stimulus, it is frequently found that other, similar stimuli are also effective and will also elicit the response that has been learned. For example, if the pigeon has been trained to peck a red spot on the wall of a cage, the pecking response will also be evoked with a lesser frequency by an orange or a yellow spot. While the property of redness is important, it is not the exclusive controlling factor. Spots of different shapes or sizes may also be effective. In general, once a response has been reinforced in one situa-

tion, the probability that the response will occur in other, similar stimulus contexts is increased. This phenomenon, which occurs when a stimulus situation different from the one in which the learner has been trained is the occasion for the occurrence of the learned response, is known as stimulus generalization. The term "generalization" refers to making the same kinds of responses to different stimuli; as such, it involves learning common elements and disregarding differences so that responses are made to new situations that are in some way similar to the situation in which previous learning occurred. Whether or not generalization is desirable or undesirable, appropriate or inappropriate, depends upon the particular task being acquired and its purpose.

The process whereby behavior is brought under the control of a single stimulus property or a special combination of properties, while not being under the control of other properties, Skinner refers to as "abstraction." For example, by reinforcing responses to a circular red spot while extinguishing responses to circular spots of all other colors, the red spot gains exclusive control over the behavior. In this case, the response is under the control of circular red spots but not under the control of the property of "redness" alone. The latter is achieved by reinforcing responses to many objects all of which are red but which differ widely in their other properties. With some degree of precision then, the organism responds to the property of redness; in this sense, an abstract response has been achieved which controls behavior in most circumstances except instances where other properties in the situation also exert control of similar strength.

Abstractions are best seen in the development of verbal behavior. The contingencies provided by other persons in the verbal community are examples of how abstractions are achieved and change and of how increasingly subtle properties of the external world come to control behavior. Abstraction is not a form of action on the part of the organism but simply a tightening of the control exercised by the properties of stimuli. Skinner (1957) writes:

Suppose, for example, that the community repeatedly reinforces a verbal response in the presence of a small red pyramid. Provided there is no interference from other behavior, the response will henceforth be evoked with varying degrees of probability by any red stimulus, any small stimulus, and any pyramidal stimulus. It is unlikely, however, that the community will also reinforce the response whenever it is made to any one of these fragmentary properties of the stimulus occurring in other combinations. If the response is to be of practical use, it must be pinned down to perhaps one property—let us say shape. The community refrains from reinforcing responses emitted in the presence of red or small objects which are not pyramidal. It continues to reinforce the response, however, whenever any pyramid is present regardless of color, size, or other property. The resulting verbal operant would traditionally be called "the name of the shape of a pyramid" and classified as *abstract*. (p. 107)

2.7 Complexity and Optimality

In his original work with animals, Skinner showed that as a result of the arrangement of contingencies of reinforcement that are manipulated in precise ways, animals could be trained to perform complex novel sequences of behavior. The various stimuli that can come to control behavior and the various stimulus-response units that can be built up produce complex behavior never before in the animal's repertoire and very different from the behavior originally emitted. The behavior-environment arrangements employed are similar to those previously described in this chapter. In his early work, *The Behavior of Organisms* (1938), Skinner describes a *tour de force* in which a rat had been trained to execute an elaborate series of behaviors—trained much in the way that an animal trainer does, through a series of approximations and appropriate arrangements of environmental contingencies. A popular account of the accomplishments of this rat, with photographs, appeared in *Life* magazine, May 31, 1937.

In Skinner's system, it is the modification of behavior to produce complex performances that has been emphasized as a framework for practical implementation. In addition to the concern with the production of behavior, there is a concern with optimality, that is, the best arrangement of training conditions that permits a subject to achieve some criterion of performance. In contrast to the design of experiments that investigate the effects of contrasting conditions in order to prove or disprove a hypothesis or theory of behavior, Skinner's experimental tactic has been to introduce environmental conditions that achieve a certain kind of behavior and that maintain the organism's "motivation" for performing it. Once certain environmental contingencies that influence behavior are investigated, then work is carried out to produce a resultant performance efficiently and skillfully. The work reported on errorless learning, for example, is one illustration of the concern for designing environmental conditions that can influence performance in optimal ways, that is, by increasing the speed of bringing about a given change in performance and by producing it with a minimum of frustration or difficulty.

2.8 Early Counterinfluences

Prior to Skinner's formulation of his system in 1938, there was general agreement among psychologists about the "empirical law of effect" which had been studied by Thorndike and his co-workers. This generalization stated that improvement in performance was attributable to the consequences following adequate and inadequate responses. The experimental studies carried out generally opposed the traditional associationistic view that the contiguity of a stimulus and response, without the operation of reinforcement or the law of effect, was a sufficient condition for many

kinds of learning. However, in the late 1940s and early 1950s, a number of experimenters asserted that an organism can learn through processes of perceptual or cognitive organization that depend only on temporal contiguity of stimulus and response, and those who were proponents of reinforcement as a necessary condition for learning were in the position of defending themselves before these cognitive "latent learning" experimenters. Although these cognitive theorists agreed that the law of effect had an important influence on the modification of behavior, they emphasized that the learning of cognitive organization was not satisfactorily explained by the operation of the aftereffects of reinforcement (Glaser, 1971a).

As this debate was going on, reinforcement theory was gaining additional power through numerous investigations and theoretical explanations of the various aspects of reinforcement, discrimination, and generalization; and the pursuit of the mechanisms of reinforcement continued. In the chapter on learning in the 1950 issue of the *Annual Review of Psychology*, the following was concluded: "It seems unlikely at this time that the reinforcement principle will be abandoned on the basis of any one or several experiments; instead, it appears that attacks on the principle only lead to more fruitful explorations of its complications" (Melton, 1950, p. 20).

Throughout the 1950s, debate and experimentation continued on crucial issues that grew up around the differences between S-R reinforcement theory and more cognitive theories. By the beginning of the 1960s, two major developments that significantly influenced conceptions of the nature of reinforcement were apparent: the extension of research into new and increasingly complex classes of behavior and the analysis of the constraints and artificialities imposed by the limited range of experimental situations in which reinforcement had been studied (see Glaser, 1971b).

While there continued to be much debate and experimentation about the effect and interpretation of reinforcement, a new non-Skinnerian interpretation was taking shape. The conception emerging was that reinforcement, as an independent variable in behavior modification, has its principal effects on performance, and learning occurs through the perception of relations among events in the environment and storage of information about the environment—a process that had been earlier simplistically ascribed to contiguity. Reinforcement, however, can control what is attended to and perceived and also influence the selection of responses that are emitted in the form of overt behavior (Glaser, 1971a). The gist of this new interpretation was that "new associations, or habits, form simply as a result of the contiguous occurrence of stimulus and response; the function of reward is the motivational one of 'energizing' habits already learned" (Estes, 1960, p. 758). The influence of these developments will be apparent as this chapter proceeds.

These cognitive conceptions, however, were not as "conspicuous" as were Skinner's notions at the time. Skinner's exclusive emphasis upon behavioral modification through contingencies of reinforcement and his active disinterest in internal processes and cognitive structures had pragmatic, relevant, and action-oriented appeal. This appeal was similar to the pragmatism that captured the imagination of American psychologists at the turn of the century—an attraction that, as compared to cognitive theory, had a particular fitness for application to human affairs. The spirit that motivated Thorndike and others of his time seems to be analogous to the appeal of Skinner's system, which came later and accepted somewhat different doctrines.

2.9 *Extrapolation to Human Behavior*

Some concern about the extension to human enterprises of a system worked out with infrahuman organisms is expressed at the end of *The Behavior of Organisms* (1938); Skinner discusses extrapolation and writes as follows:

The reader will have noticed that almost no extension to human behavior is made or suggested. This does not mean that he is expected to be interested in the behavior of the rat for its own sake. The importance of a science of behavior derives largely from the possibility of an eventual extension to human affairs. But it is a serious, though common, mistake to allow questions of ultimate application to influence the development of a systematic science at an early stage. I think it is true that the direction of the present inquiry has been determined solely by the exigencies of the system. It would, of course, still have been possible to suggest applications to human behavior in a limited way at each step. This would probably have made for easier reading, but it would have unreasonably lengthened the book. Besides, the careful reader should be as able to make applications as the writer. The book represents nothing more than experimental analysis of a representative sample of behavior. Let him extrapolate who will.

Whether or not extrapolation is justified cannot at the present time be decided. It is possible that there are properties of human behavior which will require a different kind of treatment. But this can be ascertained only by closing in upon the problem in an orderly way and by following the customary procedures of an experimental science. We can neither assert nor deny discontinuity between the human and subhuman fields so long as we know so little about either. If, nevertheless, the author of a book of this sort is expected to hazard a guess publicly, I may say that the only differences I expect to see revealed between the behavior of rat and man (aside from enormous differences of complexity) lie in the field of verbal behavior. (pp. 441-442)

Twelve years later, in 1950, the first textbook for the beginning psychology student on Skinner's system and its principles was published by Keller and Schoenfeld. The preface to the book states:

This book is a new kind of introduction to psychology. It is different in that it represents for the first time a point of view that is coming to guide the thinking and research of an active group of psychologists in this country. The members of this group are mainly experimentalists, laboratory workers. . . . They are

unflaggingly on the lookout for fundamental principles of behavior—principles that hold true for the white rat as well as the college student, for the dog in the laboratory harness as well as the patient on the psychoanalyst's couch, for the tribal savage as well as the sophisticated product of our own culture. (p. vii)

In 1945, Skinner moved from Minnesota to Indiana and then to Harvard, accepting a permanent appointment there at the Department of Psychology in 1948. At Harvard, he proposed an introductory course in human behavior to be incorporated into the general education program for undergraduates. In 1953, the text that emerged from this course was published under the title of *Science and Human Behavior*; in it, Skinner discusses the implications of his system for humans and the conduct of human affairs. This text and other activities generated by Skinner and his co-workers marked the strong move of operant psychology into human behavior.

3. TEACHING MACHINES AND PROGRAMMED INSTRUCTION

The move to Indiana in 1945 contained evidence and artifacts of Skinner's developing encounters with human behavior; he writes in his autobiography that he took with him from Minnesota the unfinished manuscript of *Verbal Behavior*, the manuscript of *Walden Two*, the "baby box"² with his young daughter as occupant, and also a lot of apparatus for a continuing program of animal experimentation (Skinner, 1967). *Verbal Behavior* was read widely in manuscript form as a result of a course Skinner taught at Columbia in 1947 and was eventually published many years later in 1957, 23 years after it was begun. The utopian novel *Walden Two*, describing a design for community living, was published in 1948. At Indiana, he conducted a seminar on self-control, the product of which eventually appeared in the chapter entitled "Self-Control" in *Science and Human Behavior* (1953). In this book, the study and control of human behavior were considered in far-ranging fashion including psychotherapy, the design of educational techniques, and problems of the control of human behavior in law and penology, religion, industry and commerce, self-control and personal behavior, and social and cultural institutions in general.

3.1 Application of Operant Principles

In 1948, at Harvard, a major development was the establishment of a pigeon laboratory in which Skinner worked with Charles Ferster and

² The baby box was an enclosed space for the care of children in which they were provided with a controlled environment without the constrictions of blankets and nighties. Skinner reported the experience of his wife and himself with the "baby box" in an article in the *Ladies Home Journal*, and hundreds of babies began to use the commercial version called an Aircrib.

others on the investigation and establishment of complex patterns of animal behavior. Under detailed schedules of reinforcement and stimulus control and the subtle shaping of responses, many new forms of behavior were produced in the pigeon. Once a response was acquired by the pigeon and brought under appropriate stimulus control, the behavior involved could be maintained for long periods of time and could be changed as a function of environmental conditions. Small differences in reinforcement schedules generated dramatic differences in behavior and by the manipulation of schedules of reinforcement, a wide range of changes in behavior was produced. In the introduction to the book reporting this work, *Schedules of Reinforcement* (Ferster & Skinner, 1957), the authors write the following:

The experimental analysis of schedules now permits the experimenter to achieve a degree of control over the organism which is of an entirely new order. High levels of activity may be generated for long periods of time. Intermediate and low levels of activity may also be generated. Changes in level which have hitherto seemed capricious may be more readily understood. Through an application of scheduling, extremely complicated examples of behavior can be set up, and behavior can be brought under subtle and complex stimulus control. As a result, complex processes can be studied in the lower organism at a new level of rigor. (p. 3)

In 1958, a group of psychologists started the *Journal of Experimental Analysis of Behavior* devoted to the publication of experiments carried out in the context of Skinner's system of behavior.

Numerous technical applications of Skinner's work in a wide range of settings were started by Skinner, his students, and colleagues at this time. A growing interest in pharmacology led many large drug companies to set up operant laboratories for the screening of new compounds, and similar work developed in university departments. Interaction with the Department of Psychiatry was instrumental in helping to set up a laboratory for the study of the operant behavior of psychotics. Work began on the use of operant principles in the management of psychotic patients in hospital wards, in psychotherapy, and in personal mental hygiene. Work on operant behavior in children that had begun earlier increased, and application turned to the use of operant principles in nursery schools and to the clinical and therapeutic study of autistic and retarded children. The masthead of the first issue of the *Journal of Applied Behavior Analysis* published in 1968 stated that it was begun "primarily for the original publication of reports of experimental research involving applications of the analysis of behavior to problems of social importance [and] . . . technical articles relevant to . . . issues arising from behavioral application."

The application of operant psychology to the classroom seems to have ridden on the crest of the confluence of the events described above: the developing interest in human behavior, the heady experiences that resulted in the close control of the behavior of organisms in the pigeon

laboratory, and the increasing interest in technological developments and applications.

3.2 *Application to the Classroom*

Skinner (1967) writes in his autobiography that he had thought about the application of his notions to the classroom in the late 1930s looking ahead to the education of his first child, but his interest really erupted upon visiting the fourth-grade class of his second daughter in 1953.

On November 11, as a Visiting Father, I was sitting in the back of the room in an arithmetic class. Suddenly the situation seemed perfectly absurd. Here were twenty extremely valuable organisms. Through no fault of her own the teacher was violating almost everything we knew about the learning process. (p. 406)

The momentum of these influences in Skinner's world is seen directly in the address given at a conference on current trends in psychology at the University of Pittsburgh in the spring of 1954 where Skinner demonstrated a machine to teach spelling and arithmetic. The address, "The Science of Learning and the Art of Teaching" (reported in Skinner, 1954), reflects the force, spirit, and tactics of Skinner's translation of work done on the programming of animal behavior into the educational application of programmed instruction (within a year, Skinner and many others were caught up in the teaching machine movement). While this was not the article that led to the popular enthusiasm for his ideas, it is a gem of writing and its cogency and vitality were transmitted to many at the time who were concerned with the relevance of the science of psychology to education. In order to transmit to the reader the sweep of the extrapolation involved, some lengthy quotation is necessary.

On what is known about learning that can be used:

Some promising advances have recently been made in the field of learning. Special techniques have been designed to arrange what are called "contingencies of reinforcement"—the relations which prevail between behavior on the one hand and the consequences of that behavior on the other—with the result that a much more effective control of behavior has been achieved. It has long been argued that an organism learns mainly by producing changes in its environment, but it is only recently that these changes have been carefully manipulated. (p. 86)

On reinforcement:

Recent improvements in the conditions which control behavior in the field of learning are of two principal sorts. The Law of Effect has been taken seriously; we have made sure that effects *do* occur and that they occur under conditions which are optimal for producing the changes called learning. Once we have arranged the particular type of consequence called a reinforcement, our techniques permit us to shape up the behavior of an organism almost at will. It has become a routine exercise to demonstrate this in classes in elementary psychology by conditioning such an organism as a pigeon. . . . Extremely complex performances may be reached through successive stages in the shaping process, the contingencies of reinforcement being changed progressively in

the direction of the required behavior. The results are often quite dramatic. In such a demonstration one can *see* learning take place. A significant change in behavior is often obvious as the result of a single reinforcement.

A second important advance in technique permits us to maintain behavior in given states of strength for long periods of time. Reinforcements continue to be important, of course, long after an organism has learned *how* to do something, long after it has acquired behavior. They are necessary to maintain the behavior in strength. Of special interest is the effect of various schedules of intermittent reinforcement. . . . Many of these effects would be traditionally assigned to the field of motivation, although the principal operation is simply the arrangement of contingencies of reinforcement. (pp. 86-87)

Extending the system:

In all this work, the species of the organism has made surprisingly little difference. It is true that the organisms studied have all been vertebrates, but they still cover a wide range. Comparable results have been obtained with pigeons, rats, dogs, monkeys, human children, and . . . human psychotic subjects. In spite of great phylogenetic differences, all these organisms show amazingly similar properties of the learning process. It should be emphasized that this has been achieved by analyzing the effects of reinforcement and by designing techniques which manipulate reinforcement with considerable precision. Only in this way can the behavior of the individual organism be brought under such precise control. It is also important to note that through a gradual advance to complex interrelations among responses, the same degree of rigor is being extended to behavior which would usually be assigned to such fields as perception, thinking, and personality dynamics. (p. 89)

Application to education:

The advances which have recently been made in our control of the learning process suggest a thorough revision of classroom practices and, fortunately, they tell us how the revision can be brought about. This is not, of course, the first time that the results of an experimental science have been brought to bear upon the practical problems of education. The modern classroom does not, however, offer much evidence that research in the field of learning has been respected or used. This condition is no doubt partly due to the limitations of earlier research. But it has been encouraged by a too hasty conclusion that the laboratory study of learning is inherently limited because it cannot take into account the realities of the classroom. In the light of our increasing knowledge of the learning process we should, instead, insist upon dealing with those realities and forcing a substantial change in them. Education is perhaps the most important branch of scientific technology. It deeply affects the lives of all of us. We can no longer allow the exigencies of a practical situation to suppress the tremendous improvements which are within reach. The practical situation must be changed. (p. 93)

On teaching drill subjects such as arithmetic:

From this exciting prospect of an advancing science of learning, it is a great shock to turn to that branch of technology which is most directly concerned with the learning process—education. Let us consider, for example, the teaching of arithmetic in the lower grades. The school is concerned with imparting to the child a large number of responses of a special sort. The responses are all

verbal. They consist of speaking and writing certain words, figures, and signs which, to put it roughly, refer to numbers and to arithmetic operations. The first task is to shape up these responses—to get the child to pronounce and to write responses correctly, but the principal task is to bring this behavior under many sorts of stimulus control. . . . Now, how is this extremely complicated verbal repertoire set up? In the first place, what reinforcements are used? . . . Secondly, we have to ask how the contingencies of reinforcement are arranged.

A third notable shortcoming is the lack of a skillful program which moves forward through a series of progressive approximations to the final complex behavior desired. A long series of contingencies is necessary to bring the organism into the possession of mathematical behavior most efficiently. But the teacher is seldom able to reinforce at each step in such a series because she cannot deal with the pupil's responses one at a time.

Perhaps the most serious criticism of the current classroom is the relative infrequency of reinforcement. Since the pupil is usually dependent upon the teacher for being right, and since many pupils are usually dependent upon the same teacher, the total number of contingencies which may be arranged during, say, the first four years, is of the order of only a few thousand. But a very rough estimate suggests that efficient mathematical behavior at this level requires something of the order of 25,000 contingencies. . . . Perhaps 50,000 contingencies is a more conservative estimate. In this frame of reference the daily assignment in arithmetic seems pitifully meagre. (pp. 90-92)

On the use of machines:

These requirements are not excessive, but they are probably incompatible with the current realities of the classroom. In the experimental study of learning it has been found that the contingencies of reinforcement which are most efficient in controlling the organism cannot be arranged through the personal mediation of the experimenter. An organism is affected by subtle details of contingencies which are beyond the capacity of the human organism to arrange. Mechanical and electrical devices must be used. . . . We have recorded many millions of responses from a single organism during thousands of experimental hours. Personal arrangement of the contingencies and personal observation of the results are quite unthinkable. Now, the human organism is, if anything, more sensitive to precise contingencies than the other organisms we have studied. We have every reason to expect, therefore, that the most effective control of human learning will require instrumental aid. . . . If the teacher is to take advantage of recent advances in the study of learning, she must have the help of mechanical devices.

Of course the teacher has a more important function than to say right or wrong. The changes proposed would free her for the effective exercise of that function.

The important features of the device are these: Reinforcement for the right answer is immediate . . . yet each child may progress at his own rate. . . . The device makes it possible to present carefully designed material in which one problem can depend upon the answer to the preceding and where, therefore, the most efficient progress to an eventually complex repertoire can be made. Provision has been made for recording the commonest mistakes so that the tapes can be modified as experience dictates. Additional steps can be inserted where pupils tend to have trouble, and ultimately the material will reach a point at which the answers of the average child will almost always be right. (pp. 94-96)

Response to possible objections:

Some objections to the use of such devices in the classroom can easily be foreseen. The cry will be raised that the child is being treated as a mere animal and that an essentially human intellectual achievement is being analyzed in unduly mechanistic terms. Mathematical behavior is usually regarded, not as a repertoire of responses involving numbers and numerical operations, but as evidences of mathematical ability or the exercise of the power of reason. It is true that the techniques which are emerging from the experimental study of learning are not designed to "develop the mind" or to further some vague "understanding" of mathematical relationships. They are designed, on the contrary, to establish the very behaviors which are taken to be the evidences of such mental states or processes. This is only a special case of the general change which is under way in the interpretation of human affairs. An advancing science continues to offer more and more convincing alternatives to traditional formulations. The behavior in terms of which human thinking must eventually be defined is worth treating in its own right as the substantial goal of education.

There is a simple job to be done. The task can be stated in concrete terms. The necessary techniques are known. The equipment needed can easily be provided. Nothing stands in the way but cultural inertia. (p. 96)

3.3 *Some Details of Design and Implementation*

An article that Skinner wrote for a more general audience appeared four years later in 1958 in *Science*. This article together with other publications popularized Skinner's ideas and eventually resulted in the notion of teaching machines and programmed instruction becoming an applied technological enterprise—one too quickly catapulted from its nurturing origins. The field was too readily divorced from its scientific background—a relationship that could have provided the checks and balances of careful investigation that are required for application from science to practice.

The article describes the educational need and contrasts the new kind of machine and techniques with the devices described by Pressey who recognized such possibilities in the 1920s. In the new machines, the student composes a response rather than selects it from a multiple choice set of alternatives; this is necessary for the delicate process of shaping. The machine also allows the student to pass through a carefully designed sequence of steps in a prescribed order that moves the student closer to fully competent behavior; the presentation given in each step is called the "frame." A student cannot proceed until he receives feedback about his response to a frame. After making a response, the correct answer is revealed, the student compares it with his own response, and only then can he move on. Skinner (1958) compares the machine to a private tutor: "This may suggest mass production but the effect upon each student is surprisingly like that of a private tutor" (p. 971). Particularly important in this regard is that the student is kept alert and busy, is given the oppor-

tunity to obtain feedback before moving on, and is provided with assistance and prompting to enable him to come up with the right answer.

Techniques of hinting, prompting, suggesting, and so on, are presented in the article as important aspects of programming material for the machine and are derived, Skinner mentions, from an analysis of verbal behavior. The basis for this is not given in much detail but reference is made to his 1957 book, *Verbal Behavior*. The major portion of the article gives examples of the programming of verbal and symbolic material in spelling, arithmetic operations, and high school physics. Memorizing poetry and learning geographical locations, features of diagrams and pictures, and anatomical vocabulary are briefly mentioned. Techniques of programming are described by example. A "vanishing" or "fading" procedure is used for teaching certain types of verbal behavior. For example, in learning about plants and animals from specimens or pictures where verbal responses are controlled by the nonverbal stimuli, the student is first asked to report features of a fully labeled object or picture; the labels are then gradually taken away or vanished so that the student has to respond to the object when the names are incomplete or lacking. Eventually the student is asked to report the relationships involved with no picture or diagram at all.

Other techniques are described in less rigorous ways. For example, frame 24 from a program in high school physics reads as follows: "The sun is not only very bright but very hot. It is a powerful _____ source of light." The correct response is "incandescent." Skinner (1958) writes:

Unwanted responses are eliminated with special techniques. If, for example, the second sentence in frame 24 were simply "It is a(n) _____ source of light," the two "very's" would frequently lead the student to fill the blank with "strong" or a synonym thereof. This is prevented by inserting the word "powerful" to make a synonym redundant. (pp. 973-974)

In general, these rules of verbal prompting and shaping are not presented as strong principles in this article; in fact, Skinner says that this article is not the place for a systematic review of available techniques or the kind of research that may be expected to discover techniques of verbal shaping and control. But he is optimistic that these can be derived from a general analysis of verbal processes. "Further research will presumably discover other, possibly more effective, techniques. Meanwhile it must be admitted that a considerable measure of art is needed in composing a successful program" (p. 974). But whether or not instructional programming is to remain an art or become a scientific technology, the reader is reassured by the fact that there is a final authority: the student (since a program is evaluated and redesigned on the basis of trial runs). Implicit in the technique of programming is the capability for revision on the basis of feedback from student performance.

The necessity for definition of the field of knowledge is mentioned in the *Science* article. Technical terms, facts, laws, principles, and cases need to be collected and arranged in a plausible developmental order for instruction so that material can be distributed for review and the generality of concepts facilitated. The technique described is not very specific. "The material is distributed among the frames of a program to achieve an arbitrary density. In the final composition of an item, techniques for strengthening asked-for responses and transferring control from one variable to another are chosen from a list according to a given schedule in order to prevent the establishment of irrelevant verbal tendencies appropriate to a single technique" (p. 974). A further point briefly mentioned is that programmed instruction would probably change the meaning of a "grade." If machine instruction assures mastery at every stage, then a grade will be useful in showing only how far a student has gone. Standards can be set so that if a student is able to go quickly enough in the time allowed, he should be able to get an A.

4. RAPID DEVELOPMENT

The *Science* article was a staging point for a number of developments. First, as already indicated, it separated the idea of teaching machines and programmed instruction from their experimental laboratory foundations. The techniques involved were frequently disseminated in superficial forms for practical exploitation, even though the necessity for continued research was stressed. Second, on the positive side, several notions were mentioned which became keystones for the revival of certain trends in education which had been more or less dormant: the individualization of instruction, the systematic development and evaluation of instructional materials, changes in the nature of testing and grading practices, and the general conception of a discipline of instructional design.

Interest in the field accelerated rapidly on many levels. In 1960, Lumsdaine and I, with support of the Department of Audio-Visual Instruction of the National Education Association, published a source book entitled *Teaching Machines and Programmed Learning*. Under Lumsdaine's drive for complete detail and coverage, this source book turned into a volume that contained or abstracted all known work in the field from early patents on automated instructional devices including Pressey's work on self-instructional test-scoring devices, the work of Skinner and his colleagues, research and development in the military on complex training devices, and experimental and analytical articles on programmed instruction and training current at the time. This book went into multiple printings, international editions, and publication in German and Japanese. Also in 1960 the October issue of *Fortune* magazine contained an article entitled "Can People Be Taught like Pigeons?" This article

commented on the similarity between Skinner's technique and the conditioning technique that Pavlov used to train a dog to drool at the sound of a bell, described the activity of major book publishers and equipment-manufacturing industries and the great potential of the market they anticipated, and made some comparisons between machines that were less "behavioristic" than Skinner's machine insofar as they allowed the possibility for errors and branching on the basis of errors in order to get the student back on the right path. Many other magazine and newspaper articles further popularized Skinner's notions.

4.1 *An Insular Technological Enterprise*

The flurry of activity continued. Programmed and unprogrammed textbooks appeared on programmed instruction, and teaching machines of various kinds were sold to schools, industry, and the military. Soon, programmed textbooks on a wide variety of subjects appeared, being more economical to produce and conceivably as effective as simple machines. These texts appeared at all levels of education from elementary schools through professional schools in engineering and medicine.

A society sprang up called the National Society for Programmed Instruction (NSPI). Its members were to some extent scholars in the field, but to a greater extent were practitioners who learned programming principles without the discipline of the underlying science of psychology. Today in the society there is an increased membership of individuals trained in psychology and interested in industrial training, and the name has changed to the National Society for Performance and Instruction. At the time it was formed, however, the society was, in my judgment, a disservice to the field because it encouraged the insularity of programmed instruction; rather, what was needed was that it lose its unique identity and be incorporated with other trends into the general stream of a developing science of instruction and instructional psychology.

A general concern at this time was that the uncritical acceptance and too-ready use of programmed instruction, coupled with high expectations and awkward application, would lead to disappointing outcomes; the rush toward immediate practicality would pull the field away from its loose ties with the scientific study of behavior. As it turned out, things were neither as good nor as bad as expected. The words "teaching machines and programmed learning" were heard less and less, and more was heard about "behavioral science research relevant to education" and "educational and teaching technology." The then current techniques of programmed learning represented only instances of the ways in which behavioral science might contribute to educational practice (Glaser, 1965a).

A book representing a broader view of the educational application of psychological research resulted from a conference that attempted to pull together psychological work on problems of training (Glaser, 1962,

1965b). The conference was sponsored by the military services but revolved around issues of general interest to civilian education. The military had sponsored a large effort along these lines in contracts both to organizations for inservice training programs and to universities and research organizations. The explicit purpose of the volume was to present a representative account of training research that had been carried out and to examine its implications for psychological research, not only for training purposes but for education in general. The individuals who contributed chapters to this book, with one exception, did not work within the Skinnerian system of operant psychology, and they brought to bear different points of view including systems analysis and analysis of skilled performance in terms of cognitive processes.

In 1964, in a lecture given to the Royal Society in England (Skinner, 1965), Skinner felt called upon to express some concern about current trends in programmed instruction. He pointed out that a special branch of psychology, the so-called experimental analysis of behavior, had produced a technology of teaching from which one could deduce programs, schemes, and methods of instruction and that the rise of two of its products, teaching machines and programmed instruction, had been meteoric. Within a decade, hundreds of instructional programs had been published, different kinds of teaching machines offered for sale, and societies for programmed instruction founded in a dozen countries. "Unfortunately, much of the technology has lost contact with its basic science. Teaching machines are widely misunderstood" (p. 427). The first programs that were produced were being copied only in certain superficial aspects and educators were assimilating the principles involved into their earlier philosophies and into older notions with which they were familiar. Skinner went on to say that programmed instruction was not to be confused with the method that Socrates used in the *Meno* to take the slave boy through Pythagoras's theorem on doubling the square, or with the writings of Comenius, or with some of the earlier work of Pressey and Thorndike; and operant conditioning was not to be confused with the conditioned reflexes of Pavlov.

In hopes of correcting prevalent misuses, in this lecture Skinner again pointed out the extraordinarily complex behaviors that could be developed in the pigeon that could be traced to subtle features of environmental contingencies of reinforcement. Programming principles used in the animal laboratory were discussed and certain applications to education were described including training procedures used for individuals with severe behavior disorders, programmed texts for teaching handwriting, crystallography, and neuroanatomy, and general possibilities for teaching study habits and problem solving. "In the government of the future the techniques we associate with education are most likely to prevail. That is

why it is so important that this young science has begun taking its most effective step in the development of a technology of teaching" (Skinner, 1965, p. 443).

4.2 Keller: *A System of Personalized Instruction (PSI)*

A technique for instruction that grew out of Skinnerian concepts, but departed from the use of programmed materials and teaching machines, was a system described by Fred Keller, a colleague of Skinner's and a major interpreter of the principles of operant conditioning (Keller & Schoenfeld, 1950; Keller, 1954). In an initial paper read at the American Psychological Association in 1963, Keller described a complicated and somewhat imaginary "personalized" course in psychology. In 1968, in the first issue of the *Journal of Applied Behavior Analysis*, he published an invited paper entitled "Goodbye Teacher . . ." which became the basis for the "Keller Plan."

This approach to teaching has achieved wide recognition in higher education and is being used by many college instructors. In 1974, a Center for Personalized Instruction was established at Georgetown University in Washington, D.C., and the first national conference on a Personalized System of Instruction (PSI) in higher education was held (Ruskin & Bono, 1974). At this conference, it was pointed out that PSI was 10 years old, that its literature included more than 400 papers, four major conferences had been held, and the courses being tried out at the college level numbered in the thousands. In the same year, an article appeared in *Science* entitled "The Keller Plan in Science Teaching" (Kulik, Kulik, & Carmichael, 1974). This article reported a 1972 survey by Hess that located 877 PSI courses in psychology, which was estimated to be about half the total number of science courses offered in the PSI format at that time. The demand for information resulted in the publication of "The Keller Plan Handbook" (Keller & Sherman, 1974) and an associated volume of germinal papers on the personalized system of instruction (Sherman, 1974). This latter collection indicates that the plan was being used in a wide range of subjects including physics, engineering, chemistry, biology, philosophy, and library science.

In his 1968 article, Keller described the five features of his instructional plan that distinguished it from conventional teaching procedures:

- (1) The go-at-your-own-pace feature, which permits a student to move through the course at a speed commensurate with his ability and other demands upon his time.
- (2) The unit-perfection requirement for advance, which lets the student go ahead to new material only after demonstrating mastery of that which preceded.
- (3) The use of lectures and demonstrations as vehicles of motivation, rather than sources of critical information.
- (4) The related stress upon the written word in teacher-student communication; and, finally:

- (5) The use of proctors, which permits repeated testing, immediate scoring, almost unavoidable tutoring, and a marked enhancement of the personal-social aspect of the educational process. (p. 83)

Keller points out the relationship of this instructional procedure to programmed instruction. There is the same stress upon the analysis and mastery of terminal performance and on the opportunity for individualized progression and frequent reinforcement, and so on, but he indicates that the "sphere of action" is different. The principal steps of advance are not "frames" but more like conventional homework assignments or laboratory exercises. A student's response is not simply a word or a phrase, but rather a verbal explanation of understanding of a principle, formula, concept, or use of a technique. Advancing within the program depends less upon the confirmation of small steps than upon a personal interaction between a student and a peer or a tutor in which there is an interchange of course information. In this context, a program text or teaching machine is useful but not to be equated with the course itself.

Several further aspects of the course are emphasized. The units of the course are not as simple as the response words in a program text and reflect more accurately the "size" of basic conceptions in the subject matter being taught. The fact that the instructor decides what is being taught is stressed, as is the academic interaction and conversation between students and between students and proctors. Testing is very frequent, perhaps 40 or 50 times in the course of one semester, and the student looks forward to this opportunity. Together with the proctor, the student is given an opportunity to defend his answers. Throughout, there is a friendly, intellectual relationship between the pairs of individuals who work together in the course. A particular characteristic of the system is that the distribution of final grades shows a pileup of *A*s with diminishing frequency in other grades as one moves from *A* to *D* since most students, if they do not withdraw from the course, can obtain an *A* or a *B* by doing all the work required and passing the tests. Students obtain grades *C* and *D* generally because of their failure to complete all the units of reading or laboratory before going into the final examination. Early sections of a course show a relatively high frequency of incomplete and withdrawal grades apparently because students have not learned to regularize their study habits and to pace themselves in order to equalize the number of tests taken per week throughout the term.

The major assumption of the Keller Plan is that there are means for arranging a student's environment so as to produce his or her best performance. The principles employed in designing this environment reflect Skinner's assumptions about learning and about empirically determining the effects of environmental conditions in facilitating behavior.

It is assumed that most students can attain the criterion of performance set in a course rather than having high standards attained by only a very few. The target behavior of primary focus is the student's verbal behavior, written and oral, related to the subject matter of the course. The fact that there is a focus on the production of verbal behavior is not unlike, but perhaps more intense than, the usual college course.

A review of published studies reports that final-examination performance in Keller Plan courses equals, and usually exceeds, performance in lecture-type courses (Kulik et al., 1974). The outstanding effect of the plan generally described is student enthusiasm. Johnston and Pennybacker (1971) report overwhelming popularity among students despite the increased time and effort and high standards required for successful performance as compared with more traditional courses. Kulik and his colleagues (1974) write that while faculty users may describe the Keller Plan as a programmed approach to instruction and may stress its link to Skinner's behavioral control and technology, students see the plan as neither controlling nor technocratic. "Instead they like its freedom, self-determination, and personal interaction. In their appreciation of the Keller Plan, faculty members may be Skinnerian, but students apparently are not" (p. 380).

The Keller Plan is not without its critics. In a review of a course prepared by Keller on the history of psychology, Jaynes (1975) writes:

What the Keller Method in complex subjects does is to grind down the individual student's particular style of going over things. It rebukes his wish to insert his own importances. The good student will hate the intrusiveness of the designer of the program text who on every page is yanking his initiative away. . . . What is learned by the Keller Method is also much more than the material: It is an attitude of subservience and dependence that comes with the removal of the material from the interpersonal community in which it is real. And I suggest that the learned subservience will last long after the material is forgotten.

Make no mistake. The Keller Method, I think, will prove extremely successful at the lower levels of education, in simpler subjects, perhaps with less creative students. I myself could wish that in some of my early training I could have escaped some of the obnoxious teaching that glommed up a particular subject matter for me thereafter, something use of the Keller Method would have joyously avoided. (pp. 630-631)

A recent review of problems and progress (Boud, Bridge, & Wiloughby, 1975) indicates the following:

One thing is clear about PSI. Many teachers have found it a very acceptable teaching method and they have been able to use it within the constraints of their present teaching system. It is undeniably one of the most popular innovations that have been proposed in higher education in the last ten years. Hopefully, its strengths and weaknesses will be determined and the nature of the teaching environments in which it proves successful will be documented. If

PSI is to prove to be more than one more fashion in the history of teaching methods, then its proponents will need to temper their enthusiasm with unbiased empirical evidence about its practice. (p. 32)

At the present time, PSI is being pushed with the fervor of adherents to Skinner's ideas to which we have become accustomed; also, as is the usual case with programmed instruction, it is controversial. The range of its applicability to various subject-matter domains and to different kinds of students and teachers needs to be studied, and its effectiveness with respect to appropriate criteria for educational success needs to be assessed. Nevertheless, PSI is influencing instructional practices in higher education.

4.3 *Individualizing Instruction*

At other levels of schooling, ideas of programmed instruction served to revitalize an old notion in the history of education. Changes in the traditional classroom structure were suggested. Introducing programmed instruction into a classroom meant that students would proceed at their own pace and would not keep together as a group; the presentation and lecturing skills of the teacher would be deemphasized, at least to the extent that more of a tutorial function would be called for. Increased time would be spent by the student on self-study and self-management of instruction, and in general, instruction would be closely tied to the progress and accomplishment of each individual. These ideas were reminiscent of attempts by educators to "individualize" instruction. As early as 1920, the public schools of Winnetka, Illinois, began using an instructional system that became known as the Winnetka Plan (Washburne & Marland, 1963). Well-known educators of the time assembled their thinking, and reports of their efforts to design individualized settings appeared in the 1925 yearbook of the National Society for the Study of Education (Whipple, 1925) entitled *Adapting the Schools to Individual Differences*. In this volume, Carleton Washburne (1925) called attention to the fact that the increasingly widespread uses of intelligence and achievement tests emphasized the realization that children varied greatly as individuals and that "it has become palpably absurd to expect to achieve uniform results from uniform assignments, made to a class of widely differing individuals" (p. x). Some way needed to be found to adapt schools to the different individuals who attended them. Washburne called for a variety of experiments, which indeed were attempted at that time.

This old desire for adapting to individual differences was reawakened by the Zeitgeist of the 1960s. During that decade, a number of programs for individualized instruction were developed, such as Individually Prescribed Instruction for the early elementary school grades (Glaser, 1968; Lindvall & Bolvin, 1967), Project PLAN at the upper grade levels and junior high school (Flanagan, Shanner, Brudner, & Marker, 1975), the

DISTAR program of Bereiter and Engelmann for disadvantaged children (1966), the Primary Education Program (Resnick, 1967) that combined prescribed instruction with the notions of the open classroom, Individually Guided Education (Klausmeier, 1975), systems of tutorial instruction at the college level (Postlethwait, Novak, & Murray, 1969), and mastery learning (Block, 1971; Bloom, 1971). Computer-assisted instruction also was assisted by this trend.

5. BEHAVIOR MODIFICATION

The current widespread use of "behavior modification" techniques represents, in large part, application of the principles of operant conditioning. The work received major impetus from Skinner's 1953 book *Science and Human Behavior* and from earlier research on operant behavior in children and adults with various behavior deviations. Techniques developed in the laboratory were applied to institutionalized psychotics, autistic children, and normal and behavior-problem children in the nursery school. In some of this work, not only were psychologists the agents of modification but appropriate procedures were also taught to parents, teachers, and paraprofessionals. This work was primarily theapeutically oriented and developed rapidly in the field of personal therapy. The discussion on psychotherapy in *Science and Human Behavior* appears to have provided a rallying point. Skinner castigated psychotherapeutic procedures based upon mentalistic, intrapsychic explanations of inner causes and offered a reconceptualization of treatment applied to direct manifestations of behavior.

The work going on in the 1950s and 1960s was gathered together in 1965 in two volumes, *Research in Behavior Modification* by Krasner and Ullmann and *Case Studies in Behavior Modification* by Ullmann and Krasner. These volumes report studies by individuals, most of whom have since become productive leaders in the field. The case studies are clinically oriented and include sections describing work on severe behavior disturbances in individual patients in institutional settings; classic neurotic behaviors; other deviant behaviors in adults such as sexual deviations, tics, and stuttering; and deviant behaviors in children and mental defectives, the latter including work in classrooms. Ullmann and Krasner rather accurately predict the future directions for research and practice: different subject populations, extension to broader aspects of social behavior, and the development of new techniques incorporating developments outside the framework of operant conditioning such as studies of vicarious reinforcement, observational learning, and biofeedback procedures. Similar to but with somewhat more caution than that reflected in the evangelism of the teaching machine movement, these publications point out that a body of information has been investigated in the laboratory and

that it is currently time for the investigation of these techniques in life situations.

Since the early 1960s, then, with the work of Skinner as a stimulus, behavior modification, as applied to psychotherapy and deviant behavior, has grown at an extremely rapid pace. The results of more than a decade of work have led many to support the use of behavior modification approaches in a wide variety of settings and with diverse types of individuals and behaviors. A government report on behavior modification issued in 1975, stated that a great range of problems appear to be responsive to behavioral treatment; but it pointed out that for many types of problems validating data have yet to be obtained. Nevertheless, the existing evidence has been strong enough that a task force of the American Psychiatric Association concluded that behavior principles employed in the analysis and treatment of clinical phenomena have reached a state of development where they unquestionably have much to offer modern clinical and social psychiatry (Stolz, Wienckowski, & Brown, 1975).

5.1 *The General Approach*

In the midst of reports of research on behavior modification and procedural manuals, textbooks are appearing that represent a balance between research findings and practical intervention. For our purposes in this chapter, some sense of the kind of presentation and content they provide is useful to define further the general approach of behavior modification and its reliance on Skinner's framework and methods. Described below are salient aspects of a recent well-written book entitled *Behavior Modification in Applied Settings* (Kazdin, 1975). The focus of the book is upon the application of operant principles and the implementation of behavior modification techniques, emphasizing particularly the measurement and evaluation of the effectiveness of intervention programs. A variety of settings is considered, such as hospitals, schools, day-care centers, the home, and outpatient treatment.

As a starting point, the differences are presented that exist between the "traditional" and behavioral conceptions of human beings—the traditional view being categorized as an intrapsychic approach because of the role accorded to psychic or psychological forces that propel behavior. Such forces include drives, impulses, needs, motives, and personality traits. A typical version of this approach is referred to as the "medical model" because it is analogous to the view of disease adhered to in the practice of medicine. In this view, treating the manifest symptom alone is insufficient because the underlying condition that is causing the problem has yet to be remedied. The extension of the medical model to the treatment of psychological problems in psychotherapy, education, counseling, and rehabilitation has led to disappointing results.

Another intrapsychic position is "trait theory," which posits underlying

personality structures to account for behavior. Traits refer to consistent, relatively enduring ways of behaving, such as aggressiveness, impulsiveness, carelessness, dominance, and submissiveness, that distinguish one individual from another. Trait explanations are inadequate because they are only indirectly inferred from behavior and the behavior is rarely consistent from situation to situation. Trait explanations also provide little information about the origins or causal conditions that led to the development of behavior; when assessment and diagnosis are carried out using trait theories, they bear little relationship to the information on real-life behavior necessary for psychological treatment or for educational purposes.

In contrast to the intrapsychic approach, behavior modification rejects the traditional conception—of inferred motives and drives in the explanation of behavior—and emphasizes environmental, situational, and social determinants as major influences on behavior. The focus is upon behavior that has been learned or needs to be learned, and explicit attempts are made to influence and train behavior rather than alter underlying aspects of personality. Emphasis is placed upon the direct observation of behavior, as such, and on the events that precede and follow behavior and thereby control it, in the operant sense of control. Behavior modification techniques can then be applied to the manipulation of external events in the environment that maintain desirable or undesirable behaviors. Treatment is usually carried out in the context of settings in which the behavior requiring change is evident, such as in the home or school.

Principles of behavior modification are described in terms of three kinds of learning—classical respondent conditioning, operant conditioning, and observational learning. (Observational learning is discussed in a subsequent section in this chapter.) The principles of operant conditioning are described in terms of relationships between behavior and the antecedent and consequent environmental events that influence it. Examples of the use of positive and negative reinforcement, punishment, and extinction are described as they have been used in the laboratory and in applied demonstrations.

In the design of behavior modification techniques, particular emphasis is placed on the details of implementation and evaluation, and these are given much attention in Kazdin's book. A first issue is the identification and careful description of the behavior to be altered. The individual may not be performing this behavior at all or only infrequently, or it is performed too frequently and should be decreased or eliminated. The behavior is specified in clear-cut fashion so that few or no inferences are required to detect the response instances that are of interest. This careful definition of behavior permits it to be measured and assessed unambiguously. This is important because a preprogram rate of behavior is required as a baseline measure, and because assessment is required to reflect

behavior change after the program is begun. A program of training for individuals using behavior modification procedures usually involves a period of training in techniques for measuring and recording behavior so that this measurement is reliable and relevant to the objectives of the program. Following such training, a further distinguishing feature of behavioral modification is careful evaluation of the treatment or training program. Techniques are taught that can be used to demonstrate that the treatment program is indeed influencing and accounting for the change that occurs. These techniques have, to a large extent, been developed in the laboratory, have been elaborated for use in applied situations, and are an important aspect of the behavior modifier's methods. Reliable measurement on a continuous basis throughout the program is emphasized so that the causal relations between contingencies and behavioral change are obvious to the behavior modifier and to the individual involved.

Following a description of the use of the various principles of operant conditioning, concern is expressed with the continued significance and effectiveness of the program. What happens when the program is discontinued? Will the changes carry over to new settings? How can an individual be weaned from the reinforcing events employed in the training program to the naturally occurring contingencies that exist in everyday life? Such questions about response maintenance and transfer of training are appearing with increasing frequency in descriptions of behavior modification programs and are receiving increased attention after apparent neglect in the early work. Attention to this matter has indicated that continued maintenance of behavior and transfer to new settings are not automatic byproducts of a program but have to be considered directly by including specific procedures for transfer in the program. Examples of such procedures are the following: substituting "naturally occurring" reinforcers for extraneous events that may have been used to change a person's behavior initially; training parents, friends, or other individuals in a client's life environment to carry out the necessary contingencies; gradually removing or fading contingencies so that they are not so frequently required or are replaced by more natural and internal rewards; allowing the client to experience behavior in an increasing variety of situations; and transferring the program used by others to a program of self-control that can be carried out by the individual himself.

Behavior modification descriptions usually end with discussion of ethical aspects and concerns over the misuses of behavioral technology. Such considerations generally discuss the problem of "control," encompassing the issues of the purpose for which the behavior is being controlled, who exerts control, and whether behavior control entails serious restrictions on individual freedom. These issues are not new to behavior modification and are of general concern in many fields of science where research has

led to practical applications.

The fear that behavioral control necessarily means an abridgment of individual freedom is generally responded to with the statement that behavioral modification used appropriately, and with consent, increases an individual's functional competence so that his or her options for living in society are increased. Because debilitating or delimiting behaviors that restrict opportunities can be overcome by appropriate treatment or education, the individual is freer to select from alternatives that were not previously available. Behavior modification practices can increase rather than stifle individual freedom because improved levels of performance and competence are achieved beyond what is possible with current practices. Nevertheless, as is the case in many aspects of social functioning, safeguards must be available for violations of ethics, of individual freedom, and the possibility of personal harm.

5.2 Behavior Modification in "Natural" Settings and Token Economies

So far in this chapter we have seen the extension of operant analyses from rats to higher order primates, to human children and adults in the laboratory, and to behavior modification for individuals. However, in his 1953 book, *Science and Human Behavior*, Skinner also extends his notions to the behavior of people in groups and to "controlling agencies" such as government and law, religion, psychotherapy, economic control, education, and, finally, to entire cultures. Skinner's early thinking about the restructuring of cultures using behavioral technology as a guiding framework was initially considered in his 1948 publication of *Walden Two* and expressed most definitively in *Beyond Freedom and Dignity* published in 1971.

In the late 1960s, work began that led to the application of behavior modification techniques to populations in "natural settings," this term being defined as non-research-oriented, human-service institutional settings such as mental hospitals, centers for the retarded and those with special behavior problems, prisons, and public schools. Early work included the study of classroom environments for behavior-problem children (Bijou, 1965; Birnbrauer, Bijou, Wolf, & Kidder, 1965; Hewett, 1967; O'Leary & Becker, 1967).

The most elaborate study at the time was the one reported by Ayllon and Azrin (1965, 1968) in which a token economy was established as a significant aspect of the living environment of patients in a mental hospital. With expressed humility, the authors undertook this large task with only their experience in the laboratory. Their far-ranging hope was that if a general procedure based upon psychological practice and theory could be found to be effective with the great range of problems encountered in a mental hospital (albeit a relatively easily managed setting), it could have

great applicability in many contexts. Ayllon and Azrin speculate, for example, that a method of controlling the aggressive outburst of a destructive patient might have relevance for the control of criminal behavior outside of the hospital. An institutionally established procedure that could motivate a vegetative psychotic who has been hospitalized for 20 years might be appropriate for motivating a high school dropout to return to school. A procedure that motivates a withdrawn person to seek out the company of other patients might have relevance in building social habits in school children, and courses of action that would enable congenitally retarded children to function might have some message for developing improved methods for teaching normal children in more efficient ways.

The basic objective of the Ayllon-Azrin undertaking was to design a motivating environment based upon operant reinforcement theory. The building of such an environment would have to consider complex human motivation and consider the use of reinforcement theory in a complex and more naturalistic environment than had been used in the past. The problem was how to translate the various terms of reinforcement theory such as "response," "behavior," "reinforcer," and "extinction" in ways applicable to natural settings and complex human behaviors. In an introduction to their work (Ayllon & Azrin, 1968) the authors list some of the questions that illustrate well the problems confronting laboratory scientists who attempt to apply their knowledge to outside settings:

The only conclusive way of determining whether the Laws of Reinforcement and Extinction can be used as the basis of designing a complex motivating environment is, of course, to try it. Before we could attempt it, however, procedures had to be developed to answer such questions as the following:

How does one go about selecting and specifying the behaviors that are to be dealt with in a motivating environment? What behaviors should one deal with? Is it possible to simultaneously deal with complex behaviors and yet have some objective and standardized way of measuring them? How should the choice of the behavior that one elects to deal with in the motivating environment be governed by factors outside of the motivating environment?

How does one go about discovering what is reinforcing for a large group of people, knowing in advance that every individual differs in his preferences from every other individual and differs with respect to his own preferences at different times? Also, how can one discover reinforcers for individuals who are nonverbal or who will not for some other reason verbalize their preferences?

Given that one has discovered some reinforcers, how can one maximize the effectiveness of the reinforcer? How does one avoid problems of satiation? How does one avoid problems of reinforcers competing with each other? How does the individual know that the reinforcer is available?

How can one arrange for reinforcers to follow a response when immediacy of the reinforcer delivery is impracticable in an overall behavioral program?

How can one record whether a behavior has been performed properly without going through the impracticable procedure of continuously observing the patient? How should the responsibilities be divided up for the staff as well as for the patients? What kinds of procedures can be developed to assure the delivery of reinforcers without using automatic devices?

How does one teach a new behavior that was not previously in existence?
(pp. 7-8)

Ayllon and Azrin described their project in detail: the general procedures, the selection and definition of behaviors to be modified, the discovery and use of reinforcing events, procedures for maximizing effectiveness in the award and exchange of tokens, ways of assuring that response reinforcement contingencies are in effect, ways of developing behavior when it hardly exists at all, and general administrative and therapeutic considerations in an institutional setting.

The results of this work showed that the effect of a token economy was substantial in maintaining higher levels of desired performance than were attainable without the token reinforcers. When reinforcement was discontinued, performance dropped substantially and very quickly. When the reinforcement economy was reinstituted, the initial effects were rapidly reinstated. The procedure was quite successful in providing a motivating environment for all individuals in the setting regardless of individual characteristics or different psychiatric classification and was effective for individuals of a wide range of ages, education, and years of hospitalization. The type of job for which reinforcement was given did not appear to alter the efficacy of the reinforcement program. The program was effective not only in maintaining very simple performances, such as personal care and housekeeping duties, but also in maintaining full-time job performances carried out in the institution.

The primary function of the token was to bridge the delay between performance and reinforcement. It also provided unambiguous indication of approval not subject to an attendant's particular interpretation at the time of delivery. The token procedure eliminated the need to discover what reinforced the patient at the time the response occurred because individual preferences could be expressed at a later time when the token could be exchanged for a wide variety of reinforcers.

5.3 *Behavior Modification in Educational Settings and Token Reinforcement in the Classroom*

As compared with therapeutic applications, applications in educational settings have proceeded slowly, but the work in schools is accelerating. This is seen in the growing number of reports of work with school children (e.g., Hanley, 1970) and the growing number of books directed toward teachers and counselors and other educational specialists. A review of the research reported in the *Journal of Applied Behavior Analysis* from its beginning in 1968 through the first half of 1976 shows an increasingly large number of studies concerned with the use of behavior modification techniques in classroom settings. The bulk of these classroom studies, however, has focused not on learning or academic performance but on decreasing disruptive behavior and its obverse, increasing attention

to task. A nondisruptive and hard-at-work class is, of course, highly reinforcing to a teacher, and this can account, to some extent, for the emphasis on these variables by behavioral researchers who have often been faced with the task of convincing teachers that reinforcement strategies are worthwhile. But more important, these variables have also been easier to study because the same unit of behavior is observed repeatedly; therefore, measurement procedures directly analogous to those used in the operant laboratory can be applied. By contrast, learning outcomes that reflect changing knowledge and skill as performance objectives require attention to a variety of additional variables.

Behavior modification and contingency management practices in education are being described and exemplified in books to be used in schools of education, for example, *Behavior Modification Procedures for School Personnel* (Sulzer & Mayer, 1972), *Classroom Management: The Successful Use of Behavior Modification* (O'Leary & O'Leary, 1972), *Behavior Modification in Educational Settings* (Klein, Hapkiewicz, & Roden, 1973), and *Behavior Modification: Applications to Education* (Keller & Ribes-Inesta, 1974). Manuals of behavior modification practices, mostly derived from the work on operant conditioning, are now available for general use. A good example is a popular book entitled *Changing Children's Behavior* (Krumboltz & Krumboltz, 1972) that is designed, the authors point out, for anyone who feels any responsibility for the behavior of a child or a group of children. They write that the principles and examples presented in the book will be valuable not only for teachers or parents but also for "psychologists, school counselors and administrators, psychiatrists, physicians, ministers, camp counselors, recreation workers, social workers, grandparents, parole officers, baby-sitters—in short, anyone who wants to help young people learn more effective ways of behaving" (p. xv). This book is organized around a set of principles each of which is described and illustrated by a wealth of examples. These principles, as summarized by the authors, are presented in Table 1. A more detailed, technical presentation, particularly for educators, is to be found in Sulzer and Mayer (1972).

A review of the emerging work in training parents in behavior modification practices (O'Dell, 1974) reported that the relationships between parental contingencies and child behavior have been dramatic. Because of such success, parent training in behavior modification is a promising area of preventive mental health care. However, there is much work to be done to turn this promise of usefulness into reliable fact. Studies need to be carried out that compare the effectiveness of different approaches with various types of parents and with children with different kinds of problems. Little evidence is available about the generalization and maintenance of the changes brought about by parents. Like many areas of behavior modification, however, parent training is being quickly expanded to fill an apparent vacuum of need.

TABLE I

Summary of Behavior Modification Principles (Krumboltz & Krumboltz, 1972)

To strengthen new behavior

1. *Positive Reinforcement Principle*: To improve or increase a child's performance of a certain activity, arrange for an immediate reward after each correct performance.

To develop new behavior

2. *Successive Approximations Principle*: To teach a child to act in a manner in which he has seldom or never before behaved, reward successive steps to the final behavior.
3. *Modeling Principle*: To teach a child a new way of behaving, allow him to observe a prestigious person performing the desired behavior.
4. *Cueing Principle*: To teach a child to remember to act at a specific time, arrange for him to receive a cue for the correct performance just before the action is expected rather than after he has performed incorrectly.
5. *Discrimination Principle*: To teach a child to act in a particular way under one set of circumstances but not another, help him to identify the cues that differentiate the circumstances and reward him only when his action is appropriate to the cue.

To maintain new behavior

6. *Substitution Principle*: To reinforce a child with a previously ineffective reward, present it just before (or as soon as possible to) the time you present the more effective reward.
7. *Intermittent Reinforcement Principle*: To encourage a child to continue performing an established behavior with few or no rewards, gradually and intermittently decrease the frequency with which the correct behavior is rewarded.

To stop inappropriate behavior you may choose from four alternative principles

8. *Satiation Principle*: To stop a child from acting in a particular way, you may allow him to continue (or insist that he continue) performing the undesired act until he tires of it.
9. *Extinction Principle*: To stop a child from acting in a particular way, you may arrange conditions so that he receives no rewards following the undesired act.
10. *Incompatible Alternative Principle*: To stop a child from acting in a particular way, you may reward an alternative action that is inconsistent with or cannot be performed at the same time as the undesired act.
11. *Negative Reinforcement Principle*: To stop a child from acting in a particular way, you may arrange for him to terminate a mild aversive situation immediately by improving his behavior.

To modify emotional behavior

12. *Avoidance Principle*: To teach a child to avoid a certain type of situation, simultaneously present to the child the situation to be avoided (or some representation of it) and some aversive condition (or its representation).
 13. *Fear Reduction Principle*: To help a child overcome his fear of a particular situation, gradually increase his exposure to the feared situation while he is otherwise comfortable, relaxed, secure, or rewarded.
-

The field of behavior modification as it might relate to education is the theme of the 1973 yearbook of the National Society for the Study of Education (Thoresen, 1973). The book contains chapters by pioneers and leaders in the field who report on current developments and past experiences. Lovaas and Koegel (1973) point out that in the short span of 10 years, behavior modification has contributed in a major way to the education of the autistic child through demonstrated effectiveness. However, while the heavy reliance upon principles from the animal laboratory has resulted in rapid progress, further advances will depend upon new research on perceptual processes in children. Bijou (1973) also points to 10 years of gratifying results with young retarded children. He sees the problems of the future as those of implementation, such as training teachers and teachers' aides in behavior modification principles and techniques, establishing parent programs, and modifying primary-grade curricula to enable children from special classes to maintain and extend the skills and knowledge they have acquired. Krasner and Krasner (1973) agree that the evidence is strong that behavior modification applications in the classroom are useful for remediation purposes. They view the next step as research and application for regular school programs.

Ayllon and Azrin's (1968) demonstration of a token economy in an institutional setting and discussion of the principles involved was a seminal study, and the idea has been picked up in many different institutional settings since that time. The studies in classroom settings that were going on at the time of the Ayllon and Azrin work reported several techniques. One was the hierarchical ordering of behaviors to be achieved in the classroom; a second was social reinforcement through interactions between teacher and student and between students and their peers. The third technique was the use of token reinforcement programs whereby behavior was controlled through the use of token rewards such as points or tangible chips that could be eventually traded for backup reinforcers like privileges, comic books, toys, and, in some cases, an allowance of money given with the approval of parents. Token reinforcement programs are now being tried out in schools with increasing frequency, particularly in classrooms with emotionally disturbed and disruptive children (e.g., O'Leary & Becker, 1968; Tharp & Wetzel, 1969).

A recent review of this area (Krasner & Krasner, 1973) traces the development of the use of token economies and planned environments from early studies in institutional settings to classrooms and communities. In these settings, the basic elements of token-economy programs consist of the following: (a) observation of the behavior of the individuals for whom the program is intended and the consequences of this behavior in the specific situation in which it is occurring; (b) designation of specific behaviors as desirable in that situation and hence to be reinforced, for example, behaviors in a classroom such as staying in one's seat, attending

to verbal instructions, nondisruptive talking, and keeping on task; (c) determination of what environmental events may serve as reinforcers for the individuals involved, for example, candy, toys, teacher praise, opportunity for play, awarded points, or chance to go on an interesting trip; (d) instigation of a medium of exchange, the token, that connects the desired behavior to the reinforcer, for example, a tangible object such as a card or a stamp, a mark on a piece of paper, or a point scored; and (e) specification of effective exchange rules, for example, a not too inflated or deflated economic relationship between the amount of tokens an individual may earn and the cost of the reinforcers.

Despite the label of *token economy*, the tokens themselves are merely . . . a training device to help teachers and others in the classroom . . . to learn how to observe behavior and its consequences, how to use their own behavior in a reinforcing manner, how to respond contingently, and how to arrange the environment so as to maximize the possibility of the individual child receiving reinforcing stimuli at the appropriate time. (p. 355)

The rise in popularity of token reinforcement programs in classrooms was indicated in a 1971 review that raised many questions that need to be answered before the effectiveness of these programs can be determined (O'Leary & Drabman, 1971). First, this analysis points out that the causal factors that may account for the demonstrated effectiveness of many of these programs in changing academic and social behavior in very diverse child populations are a complex matter. The use of tokens and backup reinforcers is only one factor contributing to the effectiveness of a token reinforcement program. Many variables are involved that influence a particular program. Possibly the most important is the child's academic skill before entering a token program since academic deficiencies can preclude further progress and result in disruptive activity as a result of frustration. Second, it is necessary to assess the reinforcers in a child's natural environment so that backup reinforcers are not used to which the child has ready access. To a significant degree the success of a program is related to the initial responsiveness of the child to any sort of contingent attention by the teacher. A further influence on program success is the extent to which disruptive or off-target behavior is maintained by peers as the attempt is made to phase out the token program and use more naturally occurring reinforcers.

Of particular interest is the extent to which behavior that is influenced by token reinforcement programs generalizes to situations outside of the specific environment in which it is used. O'Leary and Drabman emphasize that generalization is not to be expected, but that it needs to be specifically programmed and attended to so that it does indeed occur. There are few data on the effects obtained, but on the basis of experience, techniques are now being suggested to educators that might enhance the likelihood for achieving generalization.

5.4 *Debate and Constraints*

There is much debate about the effectiveness and theoretical interpretation of token reinforcement. A 1974 article entitled "Token Rewards May Lead to Token Learning" (Levine & Fasnacht) describes studies that question the long-range value of token reward systems. Based both on theoretical considerations and on practical experience, the authors conclude that "token approaches will do more harm than good when applied in what is an increasingly promiscuous manner." In the words of the unacceptable medical model: "The cure may be worse than the disease." The current cognitive approach of attribution theory suggests that what individuals attribute the cause of their behavior to determines their future behavior. Their perceptions of why they are carrying out an activity, for pay or for joy, influence how they will value the activity. Results obtained from research on attribution indicate that giving a reward can have an effect counter to that desired; giving points, tokens, M&Ms, etc., to a child for learning a lesson can decrease the intrinsic satisfaction of the lesson. Levine and Fasnacht refer to the research that shows that behavior changes have not been maintained nor generalization demonstrated to conditions in which there are no tokens. In fact, they point out that if disruptive behavior was involved, it probably would be quickly reinstated when tokens are removed because removal of tokens constitutes an extinction operation and not a generalization paradigm. The individuals have simply been taught to earn tokens but have not learned the intrinsic value of the responses that they have performed.

These authors write that the operant view of behavior is undoubtedly of real value in certain situations and with certain individuals, particularly in cases where it is possible to have reasonable control over the environment. The quick value of token reinforcement is especially useful in cases where antisocial and inappropriate behavior has been reinforced in children, for example, children whose behavior is reinforced by angry, punitive parents. But tokens were necessary for neither the acquisition of these problem behaviors nor their maintenance, and a first search should be for the natural reinforcers of problem behaviors as a significant point of intervention. Since there is a danger that tokens will decrease the intrinsic satisfaction of activities, they should be avoided unless there is a real danger to the person or there is no alternative.

Agents of change who attempt to implement behavior modification techniques in "natural" institutional settings are finding many factors that tend to mitigate the attainment of optimally effective results (Reppucci & Saunders, 1974). Institutional constraints produced by bureaucratic procedures and administrative matters and external pressures often dictated by political, economic, or administrative considerations need to be confronted (Pincus, 1974). Other problems arise from the use of indigenous

personnel as behavior modifiers, in contrast to the usual psychologist-subject relationship; from the constraints of limited resources, manpower, and time, not encountered in the laboratory; from the perceived inflexibility of the procedures instituted; and from the necessity for the behavior modifier to make accommodations and compromises between behavioral principles and the reality of the natural setting. Reppucci and Saunders express the frustrations of many psychologists who have newly entered into the arena of application. They conclude that what must be recognized is that behavior modification at its outset was never intended to be a basis for describing, understanding, or changing natural settings and that when such theoretical limitations are confronted, the environmental design of institutional settings will be on its way to a new stage of growth.

The well-tempered government report on behavior modification (Stolz et al., 1975) attempts to put the field in perspective through a discussion of what behavior modification is, and what it is not, and through a review of major issues. The authors point out that behavior modification currently is the center of controversy and that research and demonstration are required. There is a need for research on the efficacy of behavior modification methods with alternate approaches and on delivery systems to larger numbers of persons in less restricted settings than those institutions studied so far.

The concerns that have been expressed about the use of behavior modification methods need to be, and are being, attended to. The issue of control and manipulation of others by those in power is addressed by pointing out that in many therapeutic and educational situations the task is to change the individual in some way, but, to be effective, behavior modification requires a cooperative individual. Further, with knowledge of the behavioral principles involved, countercontrol can be exercised so that individuals can respond to any attempted manipulation. Concerns such as these have stimulated a reexamination of the assumptions and ethics of all psychosocial procedures designed for individual change in many kinds of social and institutional settings. Recent legal rulings have provided significant gains in human rights and have called attention to the possible abuses of behavior modification techniques. The issues are strong enough so that Stolz et al. (1975) conclude that public debate will surely continue and that professional evaluation of behavior modification techniques and public discussion of them will be required to help prevent abuses in the use of these procedures as well as to foster public understanding and acceptance of their beneficial aspects.

5.5 *Self-regulation*

Finally, there is an area of behavior modification concerned with the self-regulation and the self-determination of an individual's overt conduct and internal experiences through the manipulation of his behavior and

the events around him. Theory, experiment, and application in the area of "self-control" have been undergoing extensive development in recent years, particularly with respect to clinical applications and personal discipline (see Kanfer & Phillips, 1970; Bandura, 1971a, 1971b; Thoresen & Mahoney, 1974). Educational applications mostly have been speculative, but the little amount of work and thinking that has been done raises some interesting questions for education and the field of operant conditioning. In a chapter entitled "Self-Control," Skinner (1953) writes: "When a man controls himself, chooses a course of action, thinks out the solution to a problem, or strives toward an increase in self-knowledge, he is *behaving*. He controls himself precisely as he would control the behavior of anyone else—through the manipulation of variables of which behavior is a function" (p. 228). In the exercise of self-control, antecedent and consequent events in the environment are manipulated by the individual in such a way as to change the probability of the behavior to be controlled. The techniques of influencing and controlling behavior are parallel to the ways in which the external environment controls individual behavior.

Skinner lists several methods of self-control and gives some common examples. First, behavior is commonly controlled through physical restraint. An individual claps his hands to his mouth to keep himself from laughing or to stifle an embarrassing remark; similarly, he jams his hands into his pockets to prevent fidgeting or nail biting. The controlling response imposes some degree of physical restraint on a response to be controlled and there are reinforcing circumstances that arise when the response has been controlled. Second, certain eliciting or discriminative stimuli may be manipulated. Bad-tasting medicine may be swallowed quickly and followed with a better-tasting substance. A box of candy may be put out of sight to avoid overeating or a person may close his eyes in order to avoid distracting scenes. Unwanted emotional reactions may be reduced or eliminated by going away for a change of scene. Third, a person may deprive or satiate himself. He may skip lunch in order to take advantage of an invitation to dinner or he may eat lunch in order not to appear too hungry before his host. A person may control his behavior by drinking a lot of water before going to a cocktail party. An individual may bite a piece of celluloid in order to avoid biting his fingernails; in this way, satiation occurs on one form of behavior by engaging in a similar form. Fourth, a person may use aversive stimulation. He may put an alarm clock across the room in order to escape this stimulus by arousing himself. People can pair bad-tasting stimuli with tobacco and alcohol or make an announcement to others that they will pay them money every time they smoke a cigarette or take a drink.

Individuals may engage in activities in which their behavior is characteristically reinforced, extinguished, or punished. For this purpose, the individual must be able to obtain or withhold reinforcement but not do so

until a particular response has been emitted. For example, he might deny himself social contacts until he finishes a particular job, or an individual concerned with reducing his weight may tighten his belt up to a point where it has a strong aversive effect.

In concluding the discussion of self-control, Skinner (1953) is quick to return to his particular deterministic stance. Techniques such as those described above do not explain why the individual puts them into effect; in actuality, they come about because of arrangements by the external community.

Little ultimate control remains with the individual. A man may spend a great deal of time designing his own life—he may choose the circumstances in which he is to live with great care, and he may manipulate his daily environment on an extensive scale. Such activity appears to exemplify a high order of self-determination. But it is also behavior, and we account for it in terms of other variables in the environment and history of the individual. It is these variables which provide the ultimate control. (p. 240)

Nevertheless, he points out that an analysis of the behavior of self-control should make it possible to teach the relevant techniques as easily as any other repertoire.

A good deal of research and practical application has been carried out on the use of self-control techniques for individual therapy and counseling (see Kanfer & Phillips, 1970). Textbooks and popular manuals are available on the use of self-control techniques for the improvement of work and self-study habits, the elimination of unwanted behaviors such as smoking and drug taking, and the establishment of regimens for dieting and weight control.

A book by Thoresen and Mahoney on the subject, *Behavioral Self-Control* (1974), distinguishes two general strategies of self-controlling behavior that are derived directly from laboratory research on operant conditioning—environmental planning and behavioral programming. The first, environmental planning or discriminative stimulus control, involves changes in relevant situational factors *prior* to the performance of the target behavior that is to be instituted. An individual trying to reform his smoking habits might leave the house in the morning without sufficient change to buy cigarettes or inform friends that he is trying to give up smoking and they should criticize him if he does smoke. Essentially, this strategy involves the association of undesired behavior with stimuli that gradually can be reduced in frequency. A person trying to improve his study habits can remove distracting influences and objects from a room and design an environment in which only studying takes place; simultaneously, desired responses are linked to other stimuli. The second general strategy, behavioral programming or self-presented consequences, involves self-administered consequences *following* the occurrence of a desired target behavior, for example, rewarding oneself with a

special purchase after completing a task. With this strategy, individuals change their environment following rather than preceding the behavior to be controlled. Individuals evaluate their own behavior in this case. A dieter or reforming smoker engages in self-praise or self-criticism and, as a result, presents himself with tangible consequences. Self-controlling one's behavior requires observing and recording the response to be controlled similar to the way it has been done in laboratory research, for example, charting one's weight and administering appropriately contingent rewards or punishments.

It is conceivable that techniques of self-regulation can become part of school programs in ways that assist individuals in self-management and problem solving. But before discussion of this topic can be completed, it is necessary to consider theory and experimentation outside of Skinner's system of behavior. This relatively recent work returns to the study of the internal cognitive processes of humans with new theoretical orientations and objective techniques. The processes of self-regulation are being studied in this context with strong implications for new views about the applications of behavioral science to human affairs and about the kinds of environments that can be designed for the improvement of human learning and performance. Very briefly, we now turn to an examination of some of these counterinfluences that are of particular relevance to the topic of this chapter.

6. COUNTERINFLUENCES

6.1 *Social Learning Theory*

A major scholarly text on the principles of behavior modification has been written by Bandura (1969). This book, written with a clinical psychology emphasis, espouses a systematic position called "social learning theory" (Bandura, 1971a) that has roots in prior formulations by Miller and Dollard (1941) and Rotter (1954). Bandura's formulation has been adopted in the description of personality development (Mischel, 1968, 1971) and applied to work on counseling (Krumboltz & Thoresen, 1969). Social learning theory incorporates findings about reinforcement contingencies into a cognitive framework. Social learning theorists accept many of the operational principles of reinforcement and shaping from the work of operant conditioning but interpret the role of reinforcement and learning differently by placing emphasis on the cognitive symbolic functioning of human beings in acquiring new behavior and in regulating their performance. In introducing social learning theory, Bandura (1971a) writes:

The extreme behavioristic position . . . in a vigorous effort to eschew spurious inner causes . . . neglected determinants of man's behavior arising from his cognitive functioning. Man is a thinking organism possessing capabilities that

provide him with some power of self-direction. To the extent that traditional behavioral theories could be faulted, it was for providing an incomplete rather than an inaccurate account of human behavior.

In the social learning view, man is neither driven by inner forces nor buffeted helplessly by environmental influences. Rather, psychological functioning is best understood in terms of a continuous reciprocal interaction between behavior and its controlling conditions. . . . Social learning theory . . . places special emphasis on the important roles played by vicarious, symbolic, and self-regulatory processes, which receive relatively little attention even in the most contemporary theories of learning. These differences in governing processes carry certain implications for the way one views the causes of human behavior. (p. 2)

While Skinnerian theory depicts behavior as the product of directly experienced response consequences, social learning theory emphasizes the fact that humans have the capacity to learn by observation in acquiring behavior without having to build it up by gradual direct trial and error. The learning phenomena described by Skinner that result from direct experience and shaping can also occur vicariously through observation of other people's behavior and its consequences. Complex performances and emotional responses are learned observationally by witnessing the behavior of others being reinforced, punished, or ignored. People represent these observations symbolically and later use such representations to guide their actions; as a result, it is not always necessary to enact the various alternatives since they can foresee probable consequences of their actions and alter their behavior accordingly. In addition, humans are capable of self-regulation in situations that permit it to occur. They can manage the environment around them and produce consequences for their own actions in order to control their own behavior to some degree. These cognitive and self-regulative influences are important in determining how behavior is acquired and how it is expressed. Such influences upon behavior must be considered in addition to the direct influence of the external environment emphasized by Skinner.

Through observational learning or "modeling," new responses are acquired by imitating or simply observing the behavior of another individual. In contrast to the acquisition of new responses through differential reinforcement of successive approximations—the accepted behavioral formulation of Skinner—learning through observation need not involve direct reinforcement of the learner and need not, in fact, involve overt performance of the new behavior until it is performed in a situation, sometimes considerably delayed from its initial acquisition. In his work, Bandura has considered various theoretical conceptions of how modeling works and has reported a range of laboratory and applied studies demonstrating its effectiveness in facilitating the performance of behaviors already in the individual's repertoire, inhibiting inappropriate responses, and developing novel responses. Observation of the model's performance

leads, according to Bandura's formulation, to the establishment of "representational schemes"—imagery and verbal encoding—that on a subsequent occasion provide "a basis for self-instruction regarding the manner in which component responses must be combined and sequenced to produce new patterns of behavior" (Bandura, 1969, p. 141). Studies have shown that when several different models are observed, a generalized class of behaviors rather than a specific, direct imitation of a model can be learned.

In Bandura's theory, as in current cognitive-psychological theory, a distinction is made between learning and performance. Learning is assumed to occur as the model is observed and symbolic mediators are generated. The extent of learning depends on the learner's level of attention to the model and on the degree to which the model's performance makes the characteristics of the behavior to be learned highly discriminable. In learning, reinforcement plays an indirect role, largely through the activation of appropriate attentional and mediational processes. Performance, on the other hand, is governed by the laws of reinforcement; the individual is more likely to perform the new behavior when he is reinforced for doing so or when he expects to be so reinforced. This expectation can be fostered through observation of reinforcement contingencies applied to the model's performance. Vicarious reinforcement of this kind is an important factor in observational learning.

Thus, modeling theory does not deny the importance of reinforcement variables in the modification of behavior. But it does direct attention to other ways of learning, and it suggests strongly that the learning process can be considerably shortened by making use of the human being's capacity to learn through symbolic processes, that is, from being shown what to do and observing the effects of performance by others. Although behavior can be shaped into new patterns by rewarding or punishing consequences, learning generally would be exceedingly laborious and hazardous if it proceeded solely in this way. In contrast, most of the behaviors that people display are learned, either deliberately or inadvertently, through the influence of an example and by observation of and demonstrations by a model performer.

Most research on modeling has studied the acquisition of social rather than cognitive behaviors. The major exceptions lie in the development of language. One example of the use of modeling, in conjunction with a carefully designed differential reinforcement program, is the work of Lovaas and his associates in training language and related conceptual behaviors in autistic children (Lovaas, Berberich, Perloff, & Schaeffer, 1966; Lovaas & Koegel, 1973). Cazden (1968) has demonstrated the usefulness of a general modeling approach in language instruction for disadvantaged children. Several other studies have demonstrated the role

of models in modifying syntactic productions of children (Bandura & Harris, 1966; Liebert, Hanratty, & Hill, 1969; Rosenthal & Whitebrook, 1970). A study by Sullivan (1967) found that the viewing of films of adults conserving liquid and substance improved conservation performance of first graders, and modeling procedures have been used to modify children's information-seeking strategies (Rosenthal, Zimmerman, & Durning, 1970). Modeling principles have also been used in the development of programs to train teachers in skills that combine social and cognitive behaviors (McDonald & Allen, 1967; McDonald, 1973; Thoresen & Hosford, 1973). McDonald & Allen (1967) showed that provision of videotaped models of effective teaching behaviors enhanced acquisition of those behaviors when the trainees viewed them along with an experimenter who pointed out both the appropriate behaviors of the model and the occasion (student responses) suitable for those behaviors. The concept of modeling is being extended to learning from verbal instructions, that is, from being *told* what to do. Bandura calls this instruction "verbal models," as opposed to the more widely studied "behavioral models" (films, puppets, or live performers of the behavior). Instructions, he points out, have both an instigational function—getting performance started and perhaps setting the subject to processing information in a particular way—and a modeling function—describing the action required; however, little work along these lines has been done up to this time (Bandura, 1971a).

The focus of social learning theory and observational learning on cognitive functions highlights the difference between these interpretations and those of operant analyses. Bandura incorporates cognitive processes (coding, imagery, symbolic representations, problem solving) into his theoretical accounts of observational learning and the planning and evaluation of one's own actions that accompany it. His experimental work illustrates how such cognitive processes are involved in attenuating the influence of reinforcement contingencies, in controlling one's own actions, in thinking out and evaluating alternative actions, and in supporting or altering one's self-concept. Social learning theory thus provides a framework that is compatible with recent information-processing theories of memory, imagery, and problem solving. And these cognitive theories, for their part, can describe the details of the "representational systems" and "symbolic processes" that are employed for explanatory purposes in applications of social learning theory (Hilgard & Bower, 1975, chap. 15).

6.2 *Self-reinforcement: Personal Standards and Achievement*

The processes of reinforcement are elaborated by Bandura (1971b) in the context of observational learning and its influence on setting standards for one's behavior. Under naturalistic conditions where reinforcement occurs in a social context and individuals observe the behavior of

others and the occasions on which this behavior is rewarded, ignored, or punished, these observations provide reference standards that determine whether a particular event will serve as a reward or as a punishment. A compliment, for example, is likely to be punishing for persons who have seen similar performance by others highly acclaimed, but it is likely to be positively reinforcing when they have seen others praised less highly. Since most human behavior is not controlled by immediate external reinforcement, people regulate their actions as a result of their observations and expectations of the effects of reinforcement on others. In this context, vicarious reinforcement behavior is influenced not only by these socially observed consequences, but also by self-evaluative responses that individuals make to their own behavior. People set themselves certain performance standards and respond to their own behavior in self-rewarding and self-critical ways according to self-imposed demands. Such self-monitoring mechanisms alter and maintain behavior in the absence of external events.

The emphasis of social learning theory on human cognitive activity has further elaborated the processes involved in self-regulation as exercised by individuals through self-reinforcement processes (Bandura, 1971b). In addition to the development of criteria for evaluating the adequacy of one's performance, individuals in the absence of clear-cut criteria for standards select a basis for self-approval or self-disparagement on the basis of social comparison with others. Such criteria for self-reward or self-derogation develop on the basis of what individuals adopt as a criterion of worthy performance. This standard may be adopted on the basis of external indication of the appropriateness of a response and on the basis of the standards that are evidenced by performance models.

An individual's self-concept and self-esteem and resulting achievement and productivity are influenced by the standards of self-reinforcement that are adopted in various situations. Studies have shown (Bandura, 1971b) that individuals exposed to models favoring lenient standards of self-reinforcement were highly self-rewarding for comparatively mediocre performances. By contrast, persons who observed models adhering to stringent performance demands later displayed self-denial for identical accomplishments. In a social learning theory interpretation, these contrasting self-reactions influence the development of self-concept and related self-evaluative processes. For example, when behavior falls short of one's evaluative standards, the person judges himself negatively. On the other hand, when performances coincide with, or exceed, a person's standards, he evaluates himself positively and holds himself in high esteem. Such personal interpretations of standards for self-reinforcement can serve a motivating and reinforcing function with respect to achievement behavior since performance standards can be a major determinant

of level of productivity. "In the present interpretation, motivational effects derive not from the goals themselves but from the fact that people respond evaluatively to their own achievements and, therefore, regulate their level of effort accordingly" (p. 260).

The fact that individuals can serve as self-reinforcers and regulate their behavior on the basis of interpretations of self-produced consequences obviously takes on importance in educational settings. In evaluating a term paper or one's school work, for example, individuals may apply different standards for themselves that in turn influence their self-esteem and the outcomes of their work. The development of the ways in which individuals in school observe their own performance, set criteria for it, and provide themselves with self-reinforcement or self-disparagement is significantly influenced by their classroom experiences with their teachers and peers and observation of the modeling influences in these settings. Laboratory studies have shown, for example, that the interpersonal relationships between students and a model, in terms of personal attraction, facilitate the emulation of a model's exacting norms. In the presence of conflicting standards imposed by adult and peer models, children are predisposed toward peer modeling. Children exposed to conflicting modeling influences are inclined to reward themselves for low achievement in comparison to children who have observed an adult model consistently adhering to a high standard of self-reinforcement (Bandura, 1971a).

6.3 Covert Self-reinforcement

An individual, rather than self-administering a tangible reinforcer, can produce a symbolic, implicit consequence that can serve a reinforcing function in regulating overt behavior. Symbolic outcomes may possess reinforcing properties similar to external physical events. Studies have shown that when adults were asked to *imagine* aversive consequences, they responded in a manner similar to those who actually *experienced* the aversive consequences. The implication is that overt behavior can be partly regulated by covert self-reinforcement operations. A frontier to be explored in the investigation of self-reinforcement is the question of how symbolically produced effects can be employed as reinforcing events to control, not overt behavior, but trains of thought or covert events themselves (Homme, 1965; Thoresen & Mahoney, 1974, chap. 6). If symbolic activities obey the same psychological processes as overt behavior, then it should be possible to influence covert events when a person is trained to observe these events in himself—to self-monitor his covert performance and self-prescribe appropriate contingencies. The possibility of covert self-control was suggested by Skinner (1953) in his statement, "When we say that behavior is a function of the environment, the term 'environment' presumably means any event in the universe capable of affecting the

organism. But part of the universe is enclosed within the organism's own skin" (p. 257).

Taking a lead from Skinner and based upon the work of Bandura and others, the investigation of covert self-control is being increasingly pursued and is being facilitated by the work in cognitive psychology on imagery and other mental processes. The use of self-modification has been employed for many years in psychotherapy in techniques for the desensitization of clients from feared situations, the relaxation of tension, the extinction of negative self-thoughts, and the use of imagined negative effects to control smoking or drinking. Skills of self-reinforcement, self-control, and self-regulation have been taught to individuals in counseling and psychotherapy. In this work, an accepted practice is to shift control from the therapist or the counselor to the person himself by appropriate systematic techniques (Kanfer & Phillips, 1970). While therapy is not synonymous with education, it does involve training, and the use of self-modification techniques bears looking at with respect to educational procedures.

Of possible relevance to educational practice are studies in which self-employed imagery is used to assist in learning (Bower, 1972). An illustrative study is one in which individuals are asked to memorize a series of noun pairs such as "dog-bicycle" and to record the type of covert method they use to assist in memorization. For example, subjects might use the nouns in a sentence, create a mental picture of some action between the two nouns, or they may rehearse the pairing over and over. In the study carried out, individuals were given money or had money withdrawn for using their particular covert method for memorization, either imagery or repetition. As a result of these rewarding or punishing consequences, individuals changed their method of covert responding. It is known from experiments in human learning and memory that imagery results in better memorization, and it was found also in this study that recall of the noun pairs was much better when imagery was reinforced and reported by the subjects as the method of memorization they employed (Mahoney, Thoresen, & Danaher, 1972).

The use of covert behavior to influence cognitive performance has been studied in a training procedure in which individuals are taught to monitor "what they say to themselves" in certain situations. For example, a test-anxious student may engage in self-arousing or counterproductive statements such as "I'm going to blow this one," or "I am so nervous I'll never make it." In a formal training sequence, the individual can be taught to change his covert self-instruction to more facilitative statements, such as "If I take my time and try to relax, I'll do all right" (Meichenbaum & Cameron, 1974). Children who are impulsive in solving problems have, through similar techniques, been taught to talk to themselves in order to slow down their responses so that their performance can become more

reflective (Meichenbaum & Goodman, 1971).

In considering influences upon education in the future, it is conceivable that such concepts of self-regulation could be taught in school classes as a way of enabling the students to assist themselves in environment planning and in carrying out behavioral programs that can be used in solving intellectual problems as well as for personal problem solving. At the present time in programs of individualized instruction combined with open classroom procedures, instruction has been conducted by teaching children a variety of self-management skills in such a way that they can modify the environment for their own learning requirements (Glaser, in press).

6.4 Reinforcement and Cognition

It is evident from the research done in the context of social learning theory that vicarious reinforcement and self-reinforcement assume a prominent role in learning along with the influence of direct external reinforcement, primarily emphasized by Skinner. The operation and effects of direct external reward and punishment are attenuated by social and personal regulatory processes. The direct feedback from praise, grades in school, earned prizes or disappointments, and other external reinforcements may not influence behavior as directly as implied by behavior-modification procedures. Rather, such events may also serve as cues for the covert setting of self-standards, covert self-satisfaction or self-criticism, and the development of a particular self-concept. In complex natural situations, not as circumscribed as those generally studied in the Skinnerian framework, these factors need to be taken into account. External feedback from the environment, generally considered to be rewarding, may be personally devalued and unlikely to produce self-reward and not operate as a positive reinforcer. Attention to consistent extrinsic rewarding factors may become less important when, in the absence of extrinsic rewards, individuals can supply them for themselves. Thus, in addition to the findings of the effects of direct extrinsic reinforcement, there is now the strong evidence that the capacities of individuals can be increased to regulate their own feelings, thoughts, and actions and that circumstances in the environment can determine the relative effects of direct, vicarious, and self-produced consequences. The movement of the field of behavior modification into the arena of cognition and individual regulatory processes has been facilitated by Bandura and social learning theory.

The extension of the system to processes "beneath the skin," in keeping with modern trends in psychology, directly confronts fearsome, S-R, environmental-control aspects of the Skinnerian position. The operant conditioning framework overemphasizes one-way control of the person by the environment and neglects personal control.

The environment is only a potentiality, not a fixed property that inevitably impinges upon individuals and to which their behavior eventually adapts. Behavior partly creates the environment and the resulting environment, in turn, influences the behavior. In this two-way causal process the environment is just as influenceable as the behavior it controls. (Bandura, 1971a, p. 40)

The work on self-reinforcement and self-control has particularly made us aware that people are able to regulate their own behavior by arranging environmental conditions and administering self-reinforcing consequences. The task for education may be to teach individuals how to do this and to offer any necessary support until they can effectively guide and regulate themselves. The definition of education may be broadened to include two types of learning—knowledge of the world around us and self-knowledge about how we can best learn and acquire knowledge.

6.5 Human Factors Research and Information Processing

At the present time, modern cognitive psychology is the dominant theoretical force in psychological science as opposed to the first half of this century when behavioristic, antimentalistic S-R theories of learning were in the ascendance. In contrast, the realm of application is dominated by the particular behavioristic approach formulated by Skinner and worked on by investigators in the field of operant behavior analysis. As we have seen in the recounting in this chapter, the impact of the application of operant psychology in various forms is widespread. Skinner's work now pervades many settings where the learning and relearning of behavior are a significant phenomenon, particularly therapeutic situations, institutional environments, special education, and instructional settings at all levels of education. These applications are taking place at a fast rate, and the results obtained are beginning to present evidence with respect to the adequacies and inadequacies of the underlying scientific base. The work of the behavior modifiers has led oftentimes to impressive accomplishments, particularly in circumscribed situations and with the relatively specifiable and less complex aspects of human behavior. Certain limitations of these applications are beginning to emerge with respect to their long-range generalizable effects and their relevance to much of the internal cognitive world of complex human behavior involved in thinking, problem solving, acquiring understanding of various domains of knowledge, and personal expectations and self-development.

While basic research in operant behavior is being pursued, the present focus of activity is in applied contexts, and techniques are being made available to practitioners. In contrast, work in the laboratory and in basic theoretical developments is now burgeoning in the areas of cognitive psychology concerned with human performance, developmental psychology, and social learning. More than ever before in psychology, complex

behaviors are being investigated in rigorous scientific ways. Present-day cognitive psychology is oriented toward the complexity of human performance and is attempting to analyze it in terms of models of internal processes and how these processes are influenced by internal and external variables. However, relative to operant psychology, it is a fledgling at the present time with respect to the extent of the application of its findings and techniques to practical human endeavors.

This state of affairs is of particular interest because one of the major influences in the development of new cognitive theories, to a large extent, was motivated by applied problems. The field of application and theory was stimulated by World War II. The war led to a vast increase in research on human skills and performance, work referred to as "human engineering" or "human factors" research. Much of this work was concerned with the kind of human performance involved when individuals controlled complex man-machine systems such as aircraft and submarine systems, radar and sonar detection devices, and systems generally concerned with the detection and transmission of the information required for human and machine decision-making. Human behavior was studied in terms of its interactions with these systems, and comparisons and analogies were made between human processes and the mechanisms and behavior of mechanical and electronic systems like servomechanisms and computers. A link was forged between research in human cognitive capacities and models of these performance capabilities in terms of the hardware with which they interacted that contributed to the present-day modeling of human performance in terms of computer information-processing systems.

One of the seminal figures representative of this enterprise is Donald Broadbent who, through his distinguished career as an experimental-theoretical psychologist, was motivated by the orientation "that the test of intellectual excellence of a psychological theory, as well as its moral justification, lies in its application to concrete practical situations" (Broadbent, 1973, p. 7). Broadbent's research and theory is prototypical of the kind of theoretical concepts that emerged from practical concerns with complex human behavior and that has influenced present-day scientific cognitive psychology.

In 1958, in a book entitled *Perception and Communication*, Broadbent considered experimental work in the field of "attention" in a framework of information theory. The tasks studied reflect practical problems. Listening is studied in relation to the effects of noise in communications channels, the placement of sound sources, and the ratio of relevant to irrelevant conversation, as well as in combination with receiving visual information. The intelligibility of speech is studied over a range of noise levels and with various distortions in order to assess the ease of listening

to various communication channels, for example, radio contact of pilots with their home stations and other aircraft. The consequences of describing behavior in terms originally developed for telephone engineering are analyzed in order to identify distinctive features of human verbal behavior. Attentional processes are studied in terms of the problems of the prolonged auditory and visual vigilance tasks involved in detecting submarines by radar from the air and with respect to the effects of prolonged work such as flying an aircraft. In this book, models describing the cognitive processes involved in these complex performances are discussed in relation to theories in experimental psychology, and considerable examination is made of an emerging theory of information processing in relation to the current theorizing at that time by Hull and Skinner.

In later reflection upon new experiments growing out of this early work, Broadbent (1971) wrote in a book entitled *Decision and Stress*:

This use of applied problems gives our different specific areas another common tie; they tend to be areas requiring the man to receive and transmit information. . . . The same theoretical analyses seem to arise in tasks which superficially do not have much in common. . . . We find in our own analysis that each of them involves within the man a buffer storage, a mechanism of coding or categorization, a primary store, and so on. The common element is that of information processing, so much so that the subject matter of the book could almost be described as "psychology from the information-processing point of view." (p. 4)

The impact of these kinds of analyses of human performance is to shift emphasis away from "an unanalyzed correlation of output and input, towards a consideration of fairly complex internal processes, of observing responses, of cognitive styles, of the structure of grammatical rules, and so on. . . . New analogies and languages have made it possible to cope now with problems which psychology had put aside as too difficult a decade and more ago" (p. 2).

In 1964, the second edition of a major textbook on human factors engineering (McCormick, 1957, 1964) included a major chapter on human information processes. The chapter considers the abilities and limitations of human beings in performing various information functions such as receiving, storing, retrieving, processing, and transmitting. The words employed to discuss these topics are now the common parlance of cognitive information-processing theories of human performance, for example, the organization of information, the coding of information, and the load and speed capabilities involved in processing information.

Thus, the practical problems of skilled and complex human performance have been a shaping force in current psychological theorizing. At the present time, however, the major effort in cognitive psychology is largely concerned with attempts to develop its theoretical foundations, reminiscent of S-R theory 40 years ago. Cognitive psychology is currently

investigating complex human performances such as problem solving, language development, thinking, understanding, attention, memory, imagery, the nature of intelligence, and the problems of comprehending written and auditory discourse. These areas of human endeavor, which have been only peripheral aspects of Skinner's behavioristic psychology, now are the starting points for cognitive psychologists. Realizing quickly the lesson that has been learned—that not only might laboratory work and theory be useful for application, but that application can also be a significant test of psychological theory—cognitive psychologists (at the time of this writing) are beginning to take on educational problems.

The combination of the successes and limitations of past theories of behavior with the discoveries and potentials of new theories is what will shape the future of educational practice. It can be anticipated that prescriptions for the practical modification of behavior derived from cognitive information processing will agree closely with those of operant conditioning when one is concerned with the training of simple habits in highly restricted situations, such as has been the case in most behavior modification studies in educational settings. "But approaches deriving from the two kinds of theory may be expected to differ increasingly as we turn to problems of education and training of . . . children and adults, where multiple sources of information are the rule rather than the exception and where the individual typically has almost unlimited latitude in selecting the stimuli to which he will respond" (Estes, 1972, p. 729).

It can be further anticipated that educational practitioners will relinquish the older view that human behavior is mainly influenced directly by rewards that strengthen actions, and punishment that weakens them, and begin to design educational settings that are not restricted environments but more natural settings where human cognitive processes take on a more active role. If this takes place, then application of the techniques of operant conditioning derived from the animal laboratory will be less effective in influencing human behavior than applications that take account of the capabilities of the human organism as a processor of information (see Estes, 1972).

Representative of efforts that make a start in attempting to relate work in cognitive psychology to the solution of problems in instructional design is a recent book entitled *Cognition and Instruction* (Klahr, 1976). A chapter by Carroll (1976) examines the teaching of language skills. He points out that despite a good deal of publicity and experimentation, behavior theory has not become particularly popular with most language teachers. However, behavior theory has been the precipitating factor in the recent research attempts to teach language systems to chimpanzees. This work is beginning to throw new light on the nature of language acquisition and on certain theoretical weaknesses of the behavioristic paradigm. To its

credit, behavior theory has stimulated interesting work in the development of programmed instruction for the teaching of foreign languages and of English and has been a basis for the behavior modification techniques employed for influencing language behavior in young retarded and emotionally disturbed children. These successes have served to sort out critical elements in learning theory and have revealed some of the weaknesses of the behavioristic approach such as its inability to deal with covert events that are relevant in language learning and its failure to recognize that reinforcers affect the covert events that influence overt responses. For practical purposes, behavioristic theory provides only limited notions of the way in which responses get emitted that can serve as a basis for further learning. Carroll asks two fundamental questions: Does language learning truly take place on the basis of the restricted set of variables indicated by behavioristic theory? And when languages are acquired or modified by behavior modification techniques, does this learning result in the same kind of competence that occurs in more natural language-learning? His opinion based upon review of the experimental evidence is that the answers to both of these questions are in the negative and that there is doubt as to whether speech training conducted according to behavior modification theory would result in true language acquisition.

In offering some general prescriptions for instruction, Carroll, in the tradition of behavioral technologists, recommends careful analysis of what is to be learned. However, this analysis, with respect to language learning, is to be made in terms of information structures that correspond to meanings and communicative intentions that are manifested in overt behavior. In the preparation of instructional materials, account must be taken of what new information is presented to the learner in relation to the prior information presumed to be available. "There must be great concern with exactly what new information is presented from moment to moment in the instruction, with reference to what processing of that information is likely to be performed by the learner" (p. 21). Instead of speaking of reinforcement, we should emphasize the role that certain types of information can serve in directing the cognitive processes of the learner.

Several other chapters (Resnick, 1976; Greeno, 1976; Norman, Gentner, & Stevens, 1976; Hayes, 1976; Simon & Hayes, 1976) also emphasize the importance of task analysis, but rather than specifying "behavioral objectives," the recommendation is that cognitive objectives be developed by analyzing the psychological processes and structures in memory that are sufficient to produce the target behaviors. In this way, what becomes important in instruction is not only the answers that a student produces, but emphasis is also placed on the processes used to search for and generate the answers. Perhaps this emphasis will be one of the major

contributions of work in cognitive psychology—to provide an understanding of *process* so that educators will find it less difficult to know and to teach the kind of behavior that should be required of their students in the acquisition and use of knowledge.

The research described in *Cognition and Instruction* is typical of much work now going on. The book makes it apparent that cognitive psychologists are now preparing to study the problems of the acquisition and utilization of the intellectual skills of everyday life. For the scientists involved, an adequate theory of learning would embrace inferred covert events such as the processing of complex information, memory storage and retrieval, including the organization of knowledge structures in memory, and shifts and differences in modes of cognitive functioning that occur as children grow up and that result from the particular cultural environment to which they are exposed. At the moment, however, work focuses largely on understanding how complex cognitive tasks are performed, and to much less of an extent upon how these activities are learned—the latter being the central problem addressed by Skinner. New psychological approaches are concentrating on the question of performance and *what* is learned. For significant influence upon educational practice, the problem of learning and the acquisition of performance—the *how* of learning and teaching—must now also be considered.

7. CONCLUDING OBSERVATIONS

We are at some kind of a juncture. The systematic behavioristic analysis of learning developed in the animal laboratory is now heavily involved in the extrapolation of its concepts and techniques to practical human affairs. At the same time, cognitive theories of human performance, influenced by a concern for realistic complex human endeavors, have stimulated the present major theoretical orientation in psychology. Behavioristic applied work is being subjected to intensive tests of uses and effectiveness. Wide-scale training of practitioners and interested laymen in the techniques involved is taking place in many areas of education and human welfare. The cognitive theories of performance are focused primarily on experimental work, technique, and theory building, just as behavioristic theory was so preoccupied some years ago. But even more than behavioristic theory was at its outset, cognitive psychology is under some pressure, in the present climate of the times, to establish its practical utility. Its theories and techniques are not highly developed for the purposes of application, but a framework for an instructional psychology—a prescriptive science of educational design—may be emerging (Glaser, 1976).

The unscientific mentalism that the early behaviorists fought and that

Skinner fights now seems to be a manageable affair, and the ancient animism is at bay. The important phenomena of human thought that older mentalistic notions attempted to consider are viable subject matters again, now being investigated with new models and methodologies. By confining himself to phenomena outside the skin, Skinner has forced the identification of certain functional regularities of behavior and related successful practical applications. As a good scientist should, Skinner relentlessly presses for simplifying assumptions and for the full power of his framework, but as happens in scientific progress, constraints are being identified to account for new information that might provide deeper and more complete explanation. Corrective inputs are arising on all sides—from work with animals in the field studies of ethology and comparative psychology, from increased scientific study of child development, from increased information about language development and language comprehension, from expanded research into different natural cultural settings, from new theories of human memory and information processing, and from reexamination of the nature of human intelligence and aptitude. This overwhelming expansion of scientific investigation presents the possible hazard of overcompensation leading to too strong a rejection of what has been learned about human learning in other scientific traditions.

Skinner's notions of the "environmental space" that dictates the freedom and dignity of human actions appears to come from the adherence of his thinking to the controlled space used in operant work with animals and humans. In such situations where the environment can be specified, the behavior of an organism can be predicted and controlled with a high degree of accuracy. The success in the animal laboratory and of behavioral modification in custodial settings provides sufficient evidence. However, when we move out of these boxes and outside of highly restrictive environments, then the nature of control changes to mutual interaction between the activity of organisms and environmental conditions. The individual is now seen as receiving information from the surrounding environment and interpreting this information in order to change and to be changed by the situation. The environment is perceived, processed, and interpreted in terms of states of knowledge in an individual's memory. The prospects for new directions in education appear to be directly related to revising the concept of a restricted space. Current practical efforts in behavior modification in the classroom appear to demand restricted-space thinking. In contrast, information obtained about cognitive processes should force the design of more malleable learning situations in which the learner acts upon the environment provided, is influenced by it, and in turn changes the situation to further direct his activity.

There is little doubt that the success of operant psychology has been

limited to behavior not identified generally with the kinds of intellectual functioning and domains of knowledge to be found in the complex performances of children and adults. Effective theories of learning and instruction require knowledge of what it is that is acquired and of the components and operation of complex intellectual performance. In the classes of performance successfully worked with in behavior modification and programmed instruction, the behavior involved has been relatively easily observed and analyzed. When operant analysis has attempted to identify the nature of complex intellectual performance and the conditions under which it is learned, success has been less evident. It is unclear what the processes and component behaviors are to which the environmental reinforcing contingencies should be applied. A science leading to stronger explanations of complex human performance is necessary to identify what it is that needs to be acquired.

The pragmatic simplicity of the behavioristic, associationistic formulation that attracted the applications of Thorndike and Skinner and that is currently being transmitted to practitioners will undergo change. As complex human activity is being analyzed and better understood, we are becoming increasingly aware of the conditions that foster its acquisition and also impede its development. Current theories of child development and the influence of cultural background, the nature of human problem solving, and the understanding of structures of knowledge all appear at the moment to require constructions that cannot be subsumed under the simplifying assumptions of behavioristic theory. Cognitive psychology is striving for theoretical organization. As the theories that emerge are transferred to frameworks for optimizing human performance that are required for instructional design and theories of teaching, the complexity will be reduced. However, it can reasonably be predicted that instructional design and theories of teaching will be more complex than they are at present and, hence, require more extensively trained educators and more intricate technology for the development of instructional procedures and materials. New ways of fostering learning will emerge based upon a deeper understanding of the skills of reading and language comprehension, the skills of solving mathematical and scientific problems, and the processes that foster intelligence and aptitude. New educational practices will be discovered as a reconstruction takes place involving new knowledge of human cognition and knowledge already accumulated by behavioristic studies of learning. These new practices should lead to far fewer disabilities than presently occur in education, and to far greater accomplishments in the development of human intellect. What will be exciting in the future is the combination of the findings and fervor of Skinner's systematic study of behavior and the force of current analyses of the human mind.

Finally, it is necessary to return to the point made at the beginning of

this chapter—that education can no longer afford only the individualistic attempts of interested scientists who are motivated for some reason or other to look at problems in education. Such ventures have been fruitful for both theory and practice but have not led to the development of the “linking science” envisioned by Dewey—a structure into which knowledge can be accumulated to form a body of techniques and procedures for practical application. A prescriptive theory of instructional design is required as well as a descriptive science of human learning and performance. Such a theory of instructional design should set forth principles that specify effective ways of acquiring knowledge and mastering skills, and it should undergird a strong educational profession. A descriptive science is a necessary prerequisite, but as Skinner so forcefully has taught us, getting on with the prescriptive task should not wait, since the development of educational methods is also a strong way of assessing the limitations of our knowledge.

REFERENCES

- Anderson, G. L. (Ed.). *Learning and instruction: Forty-ninth yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1950.
- Ayllon, T., & Azrin, N. H. The measurement and reinforcement of behavior of psychotics. *Journal of Experimental Analysis of Behavior*, 1965, **8**, 357-383.
- Ayllon, T., & Azrin, N. *The token economy: A motivational system for therapy and rehabilitation*. New York: Appleton-Century-Crofts, 1968.
- Azrin, N. H., & Holz, W. C. Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application*. New York: Appleton-Century-Crofts, 1966.
- Azrin, N. H., & Lindsley, O. R. The reinforcement of cooperation between children. *Journal of Abnormal Social Psychology*, 1956, **52**, 100-102.
- Bandura, A. *Principles of behavior modification*. New York: Holt, Rinehart & Winston, 1969.
- Bandura, A. *Social learning theory*. New York: General Learning Press, 1971. (a)
- Bandura, A. Vicarious and self-reinforcement processes. In R. Glaser (Ed.), *The nature of reinforcement*. New York: Academic Press, 1971. (b)
- Bandura, A., & Harris, M. B. Modification of syntactic style. *Journal of Experimental Child Psychology*, 1966, **4**, 341-352.
- Bereiter, C., & Engelmann, S. *Teaching disadvantaged children in the preschool*. Englewood Cliffs, N.J.: Prentice-Hall, 1966.
- Bijou, S. W. A systematic approach to an experimental analysis of young children. *Child Development*, 1955, **26**, 161-168.
- Bijou, S. W. Experimental studies of child behavior, normal and deviant. In L. Krasner & L. P. Ullmann (Eds.), *Research in behavior modification: New developments and implications*. New York: Holt, Rinehart & Winston, 1965.
- Bijou, S. W. Behavior modification in teaching the retarded child. In C. E. Thoresen (Ed.), *Behavior modification in education: The seventy-second yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1973.
- Birnbrauer, J. S., Bijou, S. W., Wolf, M. W., & Kidder, J. D. Programmed instruction in the classroom. In L. P. Ullmann & L. Krasner, *Case studies in*

- behavior modification*. New York: Holt, Rinehart & Winston, 1965.
- Block, J. H. (Ed.). *Mastery learning: Theory and practice*. New York: Holt, Rinehart & Winston, 1971.
- Bloom, B. S. Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice*. New York: Holt, Rinehart & Winston, 1971.
- Boud, D. J., Bridge, W. A., & Willoughby, L. P.S.I. now—A review of progress and problems. *British Journal of Educational Technology*, 1975, **6**, 15-34.
- Bower, G. Mental imagery and associate learning. In L. Gregg (Ed.), *Cognition and learning in memory*. New York: Wiley, 1972.
- Broadbent, D. E. *Perception and communication*. New York: Pergamon Press, 1958.
- Broadbent, D. E. *Decision and stress*. New York: Academic Press, 1971.
- Broadbent, D. E. *In defense of empirical psychology*. London: Methuen, 1973.
- Carroll, J. B. Promoting language skills: The role of instruction. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Cazden, C. B. Some implications of research on language development for pre-school children. In R. D. Hess & R. M. Bear (Eds.), *Early education: Current theory, research and action*. Chicago: Aldine, 1968.
- Dewey, J. Psychology and social practice. *The Psychological Review*, 1900, **7**, 105-124.
- Estes, W. K. Learning. *Encyclopedia of educational research* (3rd ed.). New York: Macmillan, 1960.
- Estes, W. K. Reinforcement in human behavior. *American Scientist*, 1972, **60**, 723-729.
- Ferster, C. B. Positive reinforcement and behavior deficits of autistic children. *Child Development*, 1961, **32**, 437-456.
- Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.
- Flanagan, J. C., Shanner, W. M., Brudner, H. J., & Marker, R. W. An individualized instructional system: PLAN. In H. Talmage (Ed.), *Systems of individualized education*. Berkeley, Calif.: McCutchan, 1975.
- Glaser, R. (Ed.). *Training research in education* (1st ed.). Pittsburgh: University of Pittsburgh Press, 1962.
- Glaser, R. (Ed.). *Teaching machines and programmed learning, II: Data and directions*. Washington, D.C.: National Education Association, 1965. (a)
- Glaser, R. (Ed.). *Training research in education* (2nd ed.). New York: Wiley, 1965. (b)
- Glaser, R. Adapting the elementary school curriculum to individual performance. In *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton: Educational Testing Service, 1968.
- Glaser, R. Introduction. In R. Glaser (Ed.), *The nature of reinforcement*. New York: Academic Press, 1971. (a)
- Glaser, R. (Ed.). *The nature of reinforcement*. New York: Academic Press, 1971. (b)
- Glaser, R. Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, 1976, **46**, 1-24.
- Glaser, R. *Adaptive education: Individual diversity and learning*. New York: Holt, Rinehart & Winston, in press.
- Greeno, J. G. Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Hanley, E. M. Review of research involving applied behavior analysis in the classroom. *Review of Educational Research*, 1970, **40**, 597-625.

- Hayes, J. R. It's the thought that counts: New approaches to educational theory. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Henry, N. B. (Ed.). *Learning and instruction: Forty-ninth yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1950.
- Hewett, F. Educational engineering with emotionally disturbed children. *Exceptional Children*, 1967, **33**, 459-467.
- Hilgard, E. R. (Ed.). *Theories of learning and instruction: Sixty-third yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1964.
- Hilgard, E. R., & Bower, G. H. *Theories of learning* (4th ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1975.
- Homme, L. E. Perspectives in psychology: XXIV. Control of coverants, the operants of the mind. *Psychological Record*, 1965, **15**, 501-511.
- Jaynes, J. Hello, teacher . . . (Review of *The history of psychology: A personalized system of instruction* course by F. S. Keller and *Selected readings in the history of psychology: A PSI companion* by F. S. Keller). *Contemporary Psychology*, 1975, **20**, 629-631.
- Johnston, J. M., & Pennybacker, H. S. A behavioral approach to college teaching. *American Psychologist*, 1971, **26**, 219-244.
- Kanfer, F. H., & Phillips, J. S. *Learning foundations of behavior therapy*. New York: Wiley, 1970.
- Kazdin, A. E. *Behavior modification in applied settings*. Homewood, Ill.: Dorsey Press, 1975.
- Keller, F. S. *Learning: Reinforcement theory*. New York: Random House, 1954.
- Keller, F. S. "Good-bye, teacher . . ." *Journal of Applied Behavior Analysis*, 1968, **1**, 79-89.
- Keller, F. S., & Ribes-Inesta, E. (Eds.). *Behavior modification: Applications to education*. New York: Academic Press, 1974.
- Keller, F. S., & Schoenfeld, W. N. *Principles of psychology: A systematic text in the science of behavior*. New York: Appleton-Century-Crofts, 1950.
- Keller, F. S., & Sherman, J. G. *PSI: The Keller plan handbook*. Menlo Park, Calif.: Benjamin, 1974.
- Klahr, D. (Ed.). *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Klausmeier, H. J. IGE: An alternative form of schooling. In H. Talmage (Ed.), *Systems of individualized education*. Berkeley, Calif.: McCutchan, 1975.
- Klein, R. D., Hapkiewicz, W. G., & Roden, A. H. *Behavior modification in educational settings*. Springfield, Ill.: Thomas, 1973.
- Krasner, L., & Krasner, M. Token economies and other planned environments. In C. E. Thoresen (Ed.), *Behavior modification in education: The seventy-second yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1973.
- Krasner, L., & Ullmann, L. P. (Eds.). *Research in behavior modification: New developments and implications*. New York: Holt, Rinehart & Winston, 1965.
- Krumboltz, J. D., & Krumboltz, H. B. *Changing children's behavior*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.
- Krumboltz, J. D., & Thoresen, C. E. *Behavioral counseling: Cases and techniques*. New York: Holt, Rinehart & Winston, 1969.
- Kulik, J. A., Kulik, C. L., & Carmichael, K. The Keller plan in science teaching. *Science*, 1974, **183**, 379-383.

- Levine, F. M., & Fasnacht, G. Token rewards may lead to token learning. *American Psychologist*, 1974, **29**, 816-820.
- Liebert, R. M., Hanratty, M., & Hill, J. H. Effects of rule structure and training method on the adoption of a self-imposed standard. *Child Development*, 1969, **40**, 93-101.
- Lindvall, C. M., & Bolvin, J. O. Programmed instruction in the schools: An application of programming principles in "Individually Prescribed Instruction." In P. C. Lange (Ed.), *Programmed instruction: The sixty-sixth yearbook of the National Society for the Study of Education* (Pt. 2). Chicago: National Society for the Study of Education, 1967.
- Lovaas, O. I., Berberich, J. P., Perloff, B. F., & Schaeffer, B. Acquisition of imitative speech by schizophrenic children. *Science*, 1966, **151**, 705-707.
- Lovaas, O. I., & Koegel, R. L. Behavior therapy with autistic children. In C. E. Thoresen (Ed.), *Behavior modification in education: The seventy-second yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1973.
- Lumsdaine, A. A., & Glaser, R. *Teaching machines and programmed learning: A source book*. Washington, D.C.: National Education Association, 1960.
- Mahoney, M. J., Thoresen, C. E., & Danaher, B. G. Covert behavior modification: An experimental analogue. *Journal of Behavior Therapy and Experimental Psychiatry*, 1972, **3**, 7-14.
- McCormick, E. J. *Human engineering* (1st ed.). New York: McGraw-Hill, 1957.
- McCormick, E. J. *Human factors engineering* (2nd ed.). New York: McGraw-Hill, 1964.
- McDonald, F. J. Behavior modification in teacher education. In C. E. Thoresen (Ed.), *Behavior modification in education: The seventy-second yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1973.
- McDonald, F. J., & Allen, D. W. *Training effects of feedback and modeling procedures on teaching performance* (Final Report, Research Project No. OE-6-10-078, U.S. Office of Education). Stanford, Calif.: Stanford University, 1967.
- Meichenbaum, D. H., & Cameron, R. The clinical potential and pitfalls of modifying what clients say to themselves. In M. J. Mahoney & C. E. Thoresen (Eds.), *Self-control: Power to the person*. Monterey, Calif.: Brooks-Cole, 1974.
- Meichenbaum, D. H., & Goodman, J. Training impulsive children to talk to themselves: A means for developing self-control. *Journal of Abnormal Psychology*, 1971, **77**, 115-126.
- Melton, A. W. Learning. *Annual Review of Psychology*, 1950, **1**, 9-30.
- Miller, N. E., & Dollard, J. *Social learning and imitation*. New Haven, Conn.: Yale University Press, 1941.
- Mischel, W. *Personality and assessment*. New York: Wiley, 1968.
- Mischel, W. *Introduction to personality*. New York: Holt, Rinehart & Winston, 1971.
- Moore, R., & Goldiamond, I. Errorless establishment of visual discrimination using fading procedures. *Journal of Experimental Analysis of Behavior*, 1964, **7**, 269-272.
- Morse, W. H. Intermittent reinforcement. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application*. New York: Appleton-Century-Crofts, 1966.
- Munsterberg, H. *Psychology and life*. New York: Houghton, Mifflin, 1899.
- Norman, D. A., Gentner, D. R., & Stevens, A. L. Comments on learning schemata and memory representation. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

- O'Dell, S. Training parents in behavior modification: A review. *Psychological Bulletin*, 1974, **81**, 418-433.
- O'Leary, K. D., & Becker, W. C. Behavior modification of an adjustment class: A token reinforcement program. *Exceptional Children*, 1967, **33**, 637-642.
- O'Leary, K. D., & Drabman, R. Token reinforcement programs in the classroom: A review. *Psychological Bulletin*, 1971, **75**, 379-398.
- O'Leary, K. D., & O'Leary, S. G. *Classroom management: The successful use of behavior modification*. New York: Pergamon Press, 1972.
- Pincus, J. Incentives for innovation in the public schools. *Review of Educational Research*, 1974, **44**, 113-144.
- Postlethwait, S. N., Novak, J., & Murray, H. J. *The audio-tutorial approach to learning through independent study and integrated experiences* (2nd. ed.). Minneapolis, Minn.: Burgess, 1969.
- Pressey, S. L. *Psychology and the new education*. New York: Harper & Brothers, 1933.
- Reppucci, N. D., & Saunders, J. T. Social psychology of behavior modification: Problems of implementation in natural settings. *American Psychologist*, 1974, **29**, 649-660.
- Resnick, L. B. *Design of an early learning curriculum* (Publ. No. WP16). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1967.
- Resnick, L. B. Task analysis in instructional design: Some cases from mathematics. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Robinson, E. S. *Association theory to-day: An essay in systematic psychology*. New York: Century, 1932.
- Rosenthal, T. L., & Whitebrook, J. S. Incentives versus instructions in transmitting grammatical parameters with experimenter as model. *Behavior Research and Therapy*, 1970, **8**, 189-196.
- Rosenthal, T. L., Zimmerman, B. J., & Durning, K. Observationally-induced changes in children's interrogative classes. *Journal of Personality and Social Psychology*, 1970, **16**, 681-688.
- Rotter, J. B. *Social learning and clinical psychology*. Englewood Cliffs, N.J.: Prentice-Hall, 1954.
- Ruskin, R. S., & Bono, S. F. (Eds.). *Personalized instruction in higher education: Proceedings of the first national conference held by the Center for Personalized Instruction*. Washington, D.C.: Center for Personalized Instruction, 1974.
- Sherman, J. G. (Ed.). *PSI: Personalized system of instruction: 41 germinal papers*. Menlo Park, Calif.: Benjamin, 1974.
- Sidman, M., & Stoddard, L. T. The effectiveness of fading in programming a simultaneous form discrimination for retarded children. *Journal of Experimental Analysis of Behavior*, 1967, **10**, 3-15.
- Simon, H. A., & Hayes, J. R. Understanding complex task instructions. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Skinner, B. F. *The behavior of organisms: An experimental analysis*. New York: Appleton-Century, 1938.
- Skinner, B. F. *Walden two*. New York: Macmillan, 1948.
- Skinner, B. F. *Science and human behavior*. New York: Free Press, 1953.
- Skinner, B. F. The science of learning and the art of teaching. *Harvard Educational Review*, 1954, **24**, 86-97.
- Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
- Skinner, B. F. Teaching machines. *Science*, 1958, **128**, 969-977.

- Skinner, B. F. Review lecture: The technology of teaching. *Proceedings of the Royal Society*, 1965, **162**, 427-443.
- Skinner, B. F. B. F. Skinner. In E. G. Boring & G. Lindzey (Eds.), *A history of psychology in autobiography* (Vol. 5). New York: Appleton-Century-Crofts, 1967.
- Skinner, B. F. *Beyond freedom and dignity*. New York: Knopf, 1971.
- Stolz, S. B., Wienckowski, L. A., & Brown, B. S. Behavior modification: A perspective on critical issues. *American Psychologist*, 1975, **30**, 1027-1048.
- Sullivan, E. V. Acquisition of conservation of substance through film modeling techniques. In D. W. Brison & E. Sullivan (Eds.), *Recent research on the acquisition of substance*. Toronto: Ontario Institute for Studies in Education, 1967.
- Sulzer, B., & Mayer, G. R. *Behavior modification procedures for school personnel*. Hinsdale, Ill.: Dryden Press, 1972.
- Terrace, H. S. Discrimination learning with and without errors. *Journal of Experimental Analysis of Behavior*, 1963, **6**, 1-27.
- Tharp, R. G., & Wetzel, R. J. *Behavior modification in the natural environment*. New York: Academic Press, 1969.
- Thoresen, C. E. (Ed.). *Behavior modification in education: The seventy-second yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1973.
- Thoresen, C. E., & Hosford, R. E. Behavior approaches to counseling. In C. E. Thoresen (Ed.), *Behavior modification in education: The seventy-second yearbook of the National Society for the Study of Education* (Pt. 1). Chicago: National Society for the Study of Education, 1973.
- Thoresen, C. E., & Mahoney, M. J. *Behavioral self-control*. New York: Holt, Rinehart & Winston, 1974.
- Thorndike, E. L. *The psychology of arithmetic*. New York: Macmillan, 1922.
- Thorndike, E. L. *Educational psychology: The psychology of learning* (Vol. 2). New York: Columbia University, Teachers College, 1923.
- Ullmann, L. P., & Krasner, L. (Eds.). *Case studies in behavior modification*. New York: Holt, Rinehart & Winston, 1965.
- Washburne, C. W. Introduction and summary. In G. M. Whipple (Ed.), *The twenty-fourth yearbook of the National Society for the Study of Education: Adapting the schools to individual differences* (Pt. 2). Chicago: National Society for the Study of Education, 1925.
- Washburne, C. W., & Marland, S. P., Jr. *Winnetka: The history and significance of an educational experiment*. Englewood Cliffs, N.J.: Prentice-Hall, 1963.
- Whipple, G. M. (Ed.). *The twenty-fourth yearbook of the National Society for the Study of Education: Adapting the schools to individual differences* (Pt. 2). Chicago: National Society for the Study of Education, 1925.

Chapter 4

THE THEORETICAL IDEAS OF PIAGET AND EDUCATIONAL PRACTICE

by

GUY J. GROEN

McGill University

Purely negative, destructive criticism, like "refutation" or demonstration of an inconsistency does not eliminate a programme. Criticism of a programme is a long and often frustrating process and one must treat budding programmes leniently. (Lakatos, 1972, p. 179)

ANY account of the relationship between Piagetian theory and educational practice is immediately faced with a number of problems. One is the immense magnitude of the output of Piaget and his colleagues in Geneva. Another is the wide variety of applications that have been reported in the educational literature. Yet another is the fact that Piaget tends to be interpreted very differently by different people, often in a fashion that bears little resemblance to the original. It is therefore necessary to be selective yet prevent the selections from distorting one's view of the theory.

When I undertook to write this chapter, it seemed that the most reasonable approach would be to describe various applications and trace their relationship to Piagetian theory, drawing a sharp distinction between notions directly derived from Piaget, misunderstandings of Piaget, and influences from non-Piagetian sources. Upon reading through the literature, however, it became clear that this approach would not yield a coherent account of the relationship between Piagetian theory and educational practice or of the relationships of the various applications to each other. In Piagetian theory, much more than in most psychological theories, everything is related to everything else in an extremely complex fashion—to take something out in isolation is to bias the theory. Moreover, Piaget has written only a few speculative essays on education.

They do not form a part of the major thrust of his work. Contrary to the impression conveyed by much of the literature, there is no "méthode de Piaget"—no Piagetian theory of education—waiting to be applied.

As a result, this paper is not about applications per se. Rather, it is about what there is in the theory that is potentially relevant to education. Specifically, it deals with three issues. The first is the nature of the misconceptions that have arisen, especially in North America. The second is what the theory actually consists of and its scientific validity. The third is the educational implications that can be legitimately drawn.

I will begin by giving a fairly standard chronological outline of Piaget's work, and will discuss some of these misconceptions within this context. It will be seen that this standard outline is inadequate. As a result, it will be necessary to develop a new framework for concisely presenting the theory. I do this with some trepidation, since it is very difficult to prevent one's own biases from completely distorting the picture. I am not a "Piagetian." My own inclinations tend more in the direction of information-processing theories. In an effort to avoid bias, I adopt a neutral framework, based on the history and philosophy of science, and do my best to let Piaget speak for himself within this context. The approach is based on the work of the late Imre Lakatos (1972), and the resulting treatment of Piaget's theory is the most novel feature of this chapter. It will be seen that it provides a convenient basis for discussing the educational implications of the theory.

1. THE THEORY

For the purposes of this discussion, Piaget's career can be divided into three phases: the early work up to about 1930, the period from 1930 to 1957 that saw the development of the major aspects of his theory, and the period since 1957 of his major impact on North America.

1.1 *The First Phase*

Of the first phase we shall say very little. The interested reader is referred to the comprehensive reviews by Flavell (1963) and Ginsburg and Oppen (1969). The research of this period, which is reported in a series of books on language and thought (Piaget, 1923/1926), judgment and reasoning (1924/1928), reality (1926/1929), causality (1927/1930), and moral judgment (1932), is not without significance. Most important, it includes Piaget's only detailed account of affective (as opposed to cognitive) development. Indeed, these books established Piaget's international reputation and were, for many years, the only ones to have any influence in North America.

The results of this phase are quite consistent with Piaget's later work, and many of the themes developed more fully later on are present. For

example, there is a concern with the differences between the logic of children and the logic of adults. There is a division of development into stages resembling those described in the more mature work. However, the theory is much simpler. Flavell (1963, p. 270) goes so far as to claim that most of the developmental changes these books describe either are or could be interpreted in terms of a gradual replacement of egocentric thought by socialized thought in the growing child. This is more important than it might seem at first sight, because, in Piaget's work, the empirical evidence is inextricably linked to the theory. In Piagetian terms, one might say that the observations are not presented in raw form but are assimilated into the theory. Hence, any attempt to present this work in the same context as the later work demands a major effort of reinterpretation. A more critical difference concerns Piaget's empirical method. In his early work, Piaget used purely verbal tasks. In the later work, the tasks are nonverbal. Once again, this difference may not seem very important. In the context of Piagetian theory, however, it has far-reaching consequences. In the earlier work, what a child knows is assumed to be reflected in the way he or she deals with language. In the later work this assumption is abandoned and language itself occupies a very secondary role in the theory.

1.2 *The Second Phase*

The products of this period can be divided into three components. One consists of a series of books based on Piaget's observations of his own children up to the age of 4 (Piaget, 1936/1952, 1937/1954, 1945/1951). Another consists of a series of books, primarily reporting empirical observations (as usual, highly assimilated into the theory), dealing with the development of mathematical and physical concepts. The first, and most influential, deals with the concept of number (Piaget & Szeminska, 1941/1952). Two deal with geometry, topology, and the child's overall conception of space (Piaget & Inhelder, 1948/1956; Piaget, Inhelder, & Szeminska, 1948/1960). Two deal with movement and the conception of time (Piaget, 1946/1969, 1946/1970). There is also an important book contrasting logical thinking (primarily in situations that might be encountered in studying scientific subjects) of adolescents and preadolescents (Inhelder & Piaget, 1955/1958). Finally, there is a series of books dealing with more theoretical aspects. One of these is extremely well known. Piaget's *Psychology of Intelligence* (1947/1950), which represented the first comprehensive systematization of his theory, received an unusually fast and competent English translation. The other books (which appeared in 1942, 1949, 1950, and 1952) have never been translated and, with the exception of a few extracts and summaries in secondary sources such as Flavell (1963), remain generally unknown in North America.

A complete account of the work of this period is, of course, beyond the

scope of this chapter. The reader completely unfamiliar with Piaget is referred to the surveys by Flavell and by Ginsburg and Oppen previously mentioned. In order to avoid misunderstandings, however, two things must be discussed. The first is the nature of Piaget's empirical method. The second is his concept of developmental stages. Piagetian theory can be approached at a number of levels. At the most superficial level, it is a descriptive theory. Development is viewed as going through a sequence of stages: the sensorimotor stage, which ends at about age 2; the preoperational stage, which ends at about age 8; the stage of concrete operations, which ends at the age of puberty; and the stage of formal operations, which characterizes the thought of the adolescent and the mature adult. This aspect of the theory is widely known and has strong and obvious educational implications. It is therefore important to understand how it was arrived at and what it implies. Equally important, it is necessary to understand what it does not imply.

In broad terms, Piaget's empirical approach consists of placing a child in a well-defined task situation and evaluating his performance. This evaluation is accomplished by presenting subtasks or asking questions designed to probe why the child performed the task in the way he did. It has already been pointed out that, in the early work, the entire procedure is verbal, the task situation consisting of a question that the child is required to answer. In the studies of infancy, the entire procedure is, of course, nonverbal. With older children, the revised procedure of the second phase is a hybrid. The task situation is defined in terms of concrete materials that are present and the solution involves some kind of transformation or manipulation of the materials, but instructions, probes, and even the child's solution may be verbally stated. It is important to note, however, that a deliberate and systematic attempt is made to keep the results as independent of linguistic factors as possible, although the extent to which this attempt is successful is controversial. It is also important to note that the probe procedure is flexible, and the actual probe questions may vary from subject to subject. This has tended to bedevil attempts by investigators not trained in Geneva to replicate Piaget's results, since the rules (if any) by which probe items can be selected on the basis of subjects' previous behavior have never been precisely stated by Piaget or his collaborators, yet the probes are absolutely crucial in the interpretation of the results.¹

From an empirical point of view, the transitions between stages are defined in terms of the ability of children to perform specific tasks and answer the related probe questions in a satisfactory fashion. However, it is important to realize that the Piagetian stages are not monolithic entities

¹ However, a large number of Piaget's original protocols are available for anyone to examine at the Jean Piaget Archives in Geneva.

and the transitions between them are gradual rather than abrupt. For each stage, there exist a wide variety of defining tasks. Performance on these tasks is not reported in an all-or-none fashion. Instead, performance is classified into substages (primarily according to the kinds of answers given to probe items) and a child is placed in one of them. Of course, the transition can only be viewed as complete when the final substage has been attained, and it is therefore possible to classify a child as being "in" a stage on this basis.

As might be expected, there are wide individual differences in the age at which children attain a given stage. More important, perhaps, the age at which a transition occurs for a given child varies with the task. For example, an important aspect of the concrete operational stage is the notion of conservation (loosely speaking, the ability to hold a concept or property invariant under transformations of irrelevant attributes of the stimulus situation). There are many conservation tasks. Conservation of continuous quantity involves the ability to state (for the correct reasons) that the amount of liquid remains the same when it is poured from a pair of identical jars into two jars of widely disparate shapes. Conservation of substance involves the ability to state, when shown a piece of modeling clay, that the amount of clay remains the same when the shape is changed. It is well established in the Piagetian literature that conservation of substance precedes conservation of continuous quantity by several years. Similar findings are obtained with other conservation tasks.

It follows from all this that the age ranges given earlier in our brief outline of the main stages are merely rough heuristic guides. Adopting the Piagetian convention that a stage is reached on a given task at the age when 75% of the children reach the criterion level of performance, such norms merely represent the age range in which the most important tasks that characterize the stage are mastered. It also follows that it is quite inadequate to approach Piaget at this superficial level. It will be seen that the concept of stages is not really an empirical concept at all. Rather, it is a rational notion that grew from a consideration of the necessary and sufficient conditions for the existence of scientific knowledge. Since much of Piaget's recent work has been concerned with a clarification of the deeper levels of his theory, further discussion will be postponed until we have considered the third phase of Piaget's career.

1.3 *The Rediscovery of Piaget*

It has already been remarked that the work of Piaget's second phase was unnoticed in the English-speaking world for many years. Indeed, this work was carried out in relative isolation even from the rest of European psychology. Although Piaget worked with a number of collaborators, these were primarily individuals who had received their training in Geneva.

In the 1950s, this isolation came to an end. There were three main reasons for this. The first was simply the end of the Second World War and the resumption of normal scientific communications that occurred after a natural time lag. The second was that, in the early 1950s, Piaget began what appears to have been a deliberate policy of inviting established scientists and mathematicians with already well-established (and non-Piagetian) orientations to their disciplines to spend varying periods of time at Geneva.² This led to the founding, in 1955, of the Center for Genetic Epistemology which was, among other things, designed as an institution where workers from the outside could interact with Piaget and his immediate collaborators.³

The third reason was the discovery, first in England and then in North America, of the work of Piaget's second phase. This was initiated by the appearance, beginning with *The Psychology of Intelligence* in 1950, of English translations of several of his more important books. It was helped by the appearance of two articles, written by Piaget himself, in *Scientific American* (1953, 1957a) dealing with the development of mathematical and physical concepts.

By 1960 or thereabouts, enough material was available in translation that the persevering reader unfamiliar with French could get a rough idea of the work Piaget had accomplished during his second phase. However, while this material was available, it was difficult to assimilate. There were two problems. The first was that the books available, with the exception of the first book on infancy (Piaget, 1936/1952), were not self-contained. The empirical works were based on a theory that was only referred to briefly and cryptically. The only extensive⁴ theoretical work then available (Piaget, 1947/1950) did not make many specific references to the empirical works. Even worse, the empirical books themselves were interconnected in ways that were not clearly specified in the text. In other words, it was virtually impossible for the reader precisely to connect the theory with the empirical evidence or, for that matter, the various pieces of empirical evidence with each other.

The second problem was that Piaget's approach, and the issues he was dealing with, was not readily assimilable by North American psychology. At the time, North American psychology was dominated by a highly

² The first were probably the experimental psychologist Daniel Berlyne, the mathematical logician Evert Beth, and the mathematician (or computer scientist) Seymour Papert.

³ The results of some of these interactions appear in various volumes of the *Études d'Épistémologie Génétique*.

⁴ The only other book available was a very short work (Piaget, 1957b) based on a series of lectures at the University of Manchester and dealing with the more formal aspects of his theory. Curiously enough, this book is more specific, in its references to the empirical work, than the more extensive *Psychology of Intelligence*.

reductionistic, empirical approach epitomized by Hullian theory (e.g., Hull, 1943) and the six-volume survey of the field by Koch (1959-1963), with a substantive emphasis on learning as the primary process determining the course of human development, a view of maturation as unimportant, and a methodological emphasis on a close and explicit relationship between theory and experiment. With the possible exception of the work on infancy, Piaget appeared almost as a polar opposite. Many of the empirical results aroused interest, but the theory, to the extent that it was understood at all, appeared thoroughly alien. At best, it appeared as a rather bizarre effort, only loosely related to the empirical evidence, which emphasized the role of maturation and deemphasized learning.⁵

These two problems interacted. Clearly, the ideal way to solve the first would be by producing a concordance to the Piagetian literature. In 1960, however, such a concordance would have needed a considerable amount of explanatory material to be useful (a fact which may still be true today). In other words, the more useful and salable product was a guidebook rather than a concordance. Two examples have already been cited (the books by Flavell and by Ginsburg and Oppen), and many more exist today (e.g., Phillips, 1975). In order to communicate with the intended audience, however, the author of such a guide must, to some extent, present the subject matter from the point of view of the audience's presumed theoretical and methodological biases and deal with subject matter that the audience finds interesting. As a result, a guidebook to Piaget must inevitably present a version of Piaget distorted, to a greater or lesser extent, by the biases of the author and the intended audience.

If the theoretical biases of the 1950s had persisted, it is unlikely that Piaget would have achieved the prominence in North America that he has today. However, North American psychology was ripe for change. This was especially true as far as experimental psychology was concerned (and, of course, those parts of developmental and child psychology that made extensive use of experimental techniques). Broadly speaking, experimental psychology could be divided into three areas: (a) sensation and perception; (b) learning and motivation; (c) thinking, problem solving, and other forms of "complex behavior." Underlying this division was an assumption that the first two areas could somehow be studied independently of the third, that they represented more elementary forms of behavior, and that a good theory of complex behavior was impossible without good theories of learning and perception. By 1960, this assumption was being seriously questioned (e.g., Miller, Galanter, & Pribram, 1960).

The most disturbing questions concerned the nature and necessity of

⁵ In reality, Piagetian theory and Hullian theory may not be completely antagonistic. See Berlyne (1965) for an attempt to reconcile the two approaches.

theories of learning. The predominant theories were associationistic in nature, being based on the assumption that what was learned was invariably a set of associations between observable stimuli, observable responses, and hypothetical internal events. After half a century of effort, however, nobody had devised a theory of this type that could cope with complex behavior. The possibility was emerging that there might be a distressingly simple reason for this: The entire theoretical approach might be incorrect. Linguists such as Chomsky (1957) were beginning to argue that associationistic theories could not, by their very nature, explain the acquisition of grammatical rules. Psychologists such as Miller, Galanter, and Pribram (1960) were beginning to argue that the theories were inadequate to account for the acquisition of most of the more important skills underlying thinking, problem solving, and many kinds of complex human performance. The notion that a satisfactory theory of learning was somehow necessary for a satisfactory theory of thinking and problem solving was also being questioned. Direct experimental and theoretical investigations of thinking (e.g., Bruner, Goodnow, & Austin, 1956; Newell, Shaw, & Simon, 1958) seemed most successful when the research questions asked were about the strategies, hypotheses, and goals a subject used in performing a task, and the issue of how they were learned was left in abeyance. This implied that the best route to a satisfactory theory of complex human behavior might be through the analysis of performance rather than learning. Indeed, it might well be that learning, in such a context, was nothing more than a very complex form of problem solving! In any case, the notion was becoming more plausible that it might be better to examine carefully the causes of changes in performance in a task-specific fashion rather than search for general laws of learning *per se*.

For many years the vast majority of experimental psychologists had been content to ignore complex human behavior. Why was there such a sudden upsurge of interest? There were a number of reasons, among them the availability of novel techniques of theory construction and data analysis originating in mathematics, computer science, and linguistics.⁶ There was also the fact that government agencies, especially those connected with the armed forces, had been undertaking and supporting research in the areas of personnel training, human performance, and decision making. The most important reason, however, was an increased concern with the kind of psychology that might be relevant to education

⁶ The three-volume *Handbook of Mathematical Psychology* (Luce, Bush, & Galanter, 1962-1965) contains extensive discussions of most of these. Unfortunately, there is only a brief chapter on computer techniques. Many of the more important early papers on the computer simulation of cognitive processes are included in the anthology edited by Feigenbaum and Feldman (1963). There is also an interesting account of the early history of this field in the appendix to Newell and Simon (1972).

at the high-school and elementary-school levels. It had been kindled by a movement for curriculum reform, begun in the mid-1950s, that received widespread public attention and governmental support after the launching of the Russian satellite Sputnik in 1957.

The leaders of the movement were professional scientists and mathematicians. They were concerned by the fact that the science and mathematics taught in the schools bore little resemblance to what was used by professional engineers, scientists, and mathematicians. Moreover, the conventional curriculum was conveying little of the insight gained by 19th- and 20th-century mathematics and science. Not only were the schools inadequately preparing future engineers and scientists, but they also were felt to be in danger of producing a generation of scientific illiterates.

The "new curricula" that resulted (in the humanities as well as in the sciences) are so well known that they need not be described here. It suffices to say that the reforms were radical, at least in their original intent. The goal was to create curricula that would teach the average student to understand the basic conceptual structure of the discipline. To achieve this goal, it was felt that the conventional order in which subject matter was taught had to be drastically changed. Things that were commonly taught in advanced university courses had to be introduced at extremely low grade levels, the most dramatic example being the introduction of set theory into the earliest primary grades of some of the "new mathematics curricula" (e.g., Suppes, 1960/1969). To allow for the additional time required to teach these notions, those portions of the conventional curriculum that involved rote memorization or the extensive use of computational skills were either deemphasized or eliminated completely.

The implementation of these reforms raised some very basic psychological issues. The notion that a group of like-minded experts in a given discipline can get together and hammer out a curriculum that reflects their view of the conceptual structure of the discipline is quite feasible. However, teachers have to teach it and children have to learn from it. Ideally, it should be easy for both. To accomplish this, it was necessary to carefully grade or scale the difficulty of the subject matter being taught. It was also necessary to devise new teaching techniques that would somehow enhance the child's conceptual understanding. This could be done (and almost always was) by a combination of intuition and trial-and-error. However, this was a very time-consuming and expensive procedure. Also, in the absence of measuring instruments sufficiently precise to detect errors in a reliable fashion, it could only result in a curriculum with many hidden pitfalls for student and teacher alike. Were there any sound psychological principles and techniques that a curriculum developer could use in solving problems of this kind?

Unfortunately, there was no easy way of responding to these demands. The new curricula were unique in their emphases on cognitive processes and cognitive structure. Most psychologists were ill-equipped to deal with cognition directly. The educational psychology of the period primarily consisted of applications of other branches of psychology to issues raised by the older, conventional type of curriculum. Much of it was applied personality theory and social psychology. There was one area, the psychology of mental testing and individual differences, that dealt with cognitive issues, but there were serious problems here as well. There existed an elaborate correlational methodology for measuring global traits and dimensions, such as intelligence, that were correlated with the ability to perform well on complex cognitive tasks. However, nobody knew how to define these dimensions and traits more precisely. For example, there was still no agreement on a satisfactory definition of intelligence. Also, there was no satisfactory way of explaining how such traits changed. Intelligence, as measured by IQ, was an invariant quantity. Intelligence, as measured by mental age, was assumed to reflect an undefined maturational process that occurred independently of experience. In terms of this methodology, even achievement in a specific school subject was defined as a global trait. This could be reliably measured, but there was no way of saying what it was about the way the subject had been taught that determined this measure.

It was in this kind of context that the assimilation of Piagetian theory began. There was a search for new approaches. One thing that seemed to be necessary was a theoretical approach that could account for the development of complex cognitive skills in a way that was responsive to the cognitive issues raised by the new curricula. Despite the fact that Piaget had said little about education directly, his approach was the only one that seemed to satisfy these specifications at that time.

Conferences began to be held. Probably the first was the 1959 Woods Hole Conference on the improvement of science education. It was attended by Piaget's principal collaborator, Bärbel Inhelder. Another, held in 1960, dealt at length with the impact of Piagetian research on child psychology (Kessen & Kuhlman, 1962). Bärbel Inhelder was again a participant. The culmination was the 1964 Conference on Cognitive Studies and Curriculum Development (also called the Piaget Conference in the report), which was attended by Piaget himself (Ripple & Rockcastle, 1964).

Books began to appear. There were three that were particularly influential in determining how the assimilation would take place.⁷ Bruner (1960) provided a manifesto, loosely based on the proceedings of the

⁷ We speak here specifically of the United States. In England, the process of assimilation began slightly earlier and occurred in a much more specifically educational context (National Froebel Foundation, 1955; Lunzer, 1960; Lovell, 1962).

Woods Hole Conference, describing how Piagetian theory could resolve the psychological issues raised by the new curricula. Hunt (1961) dealt with the relevance of Piagetian theory to the psychology of mental testing and individual differences, with particular emphasis on the role it could play in a more satisfactory definition of intelligence. Flavell (1963) provided the first thorough guide to the Piagetian literature. Its importance cannot be overstressed. Whereas Bruner provided a popularization designed to be intelligible to the lay reader and Hunt was only concerned with those aspects relevant to his own theoretical approach, Flavell provided a complete, uncompromising, self-contained account of the entire body of Piaget's work, including much that was not available in translation. Moreover, it was presented in a form that the psychologist (especially the experimental psychologist) trained in North America could appreciate. The first part was entitled "The Theory." The second part was entitled "The Experiment." The third part was entitled "Critique" and included an evaluation of the adequacy of the experiments as tests of the theory.

By 1964, the process of rediscovering Piaget was complete. There were now the beginnings of a tradition of Piagetian research in the English-speaking world. What followed was a flood of research on Piagetian topics. No attempt will be made to characterize this work in its entirety. The interested reader is referred to the reviews by Wallace (1965) and Modgil (1974) and the collection of papers edited by Sigel and Hooper (1968). It is important to note, however, that it has differed in important respects from the Geneva tradition. There has been a concern with specific educational applications. There has been more of a tendency to invent new theories incorporating aspects of Piaget than to develop Piagetian theory *per se*. Most important, perhaps, Piagetian theory has frequently been viewed as an object to be destroyed by experimental techniques. Some methodological problems with such experiments were discussed in the preceding section. There is a more important problem with such experiments, however. Frequently, the theory under attack is a simplified subset of Piagetian theory.

Flavell's book had been too successful. Its treatment of Piaget was so thorough and well organized that it could be used as a primary source, with only occasional references to Piaget's own works to check out the details. For many English-speaking psychologists, Piagetian theory was the theory as described by Flavell rather than by Piaget himself. However, there was a change of emphasis. As Piaget puts it, in his introduction in Flavell's book, "It seems clear that Professor Flavell is more interested in the experiments than in the theory, which sometimes gives me the impression—perhaps not of having been misunderstood, but, if you will—of having been understood on certain issues more from without than from within." Flavell presented the theory rather than explained it.

Much of the exposition consisted of direct translations from a bewildering variety of Piaget's writings, with insufficient commentary or clarification. The reader attempting to reconstruct Piagetian theory from Flavell's account was supplied with all the bricks but not enough mortar. The only thing to do was to throw away some of the bricks.

1.4 *The Third Phase*

The transition to this phase was not marked by any drastic change in either theory or methodology. There was a change, nonetheless. Since 1955, there has been much more of a tendency to react directly to stimulation and criticism from the outside world, partly for reasons discussed in the preceding section and partly, no doubt, for the simple reason that there has been much more to react to. This has resulted in a major clarification of Piaget's theoretical position. An account of this deserves a section of its own. Because of this, the present section will be brief.

As with the second phase, there is a series of books by Piaget and his collaborators consisting primarily of empirical observations. However, the issues they deal with are more general. The first deals with the development of elementary logical operations, as reflected in classification and seriation tasks (Inhelder & Piaget, 1959/1964). Other books deal with mental imagery (Piaget & Inhelder, 1966/1971), memory (Piaget & Inhelder, 1968/1972), learning (Inhelder, Sinclair, & Bovet, 1974), and consciousness (Piaget, 1974/1976). There is also a considerably more theoretically oriented work dealing with perception (Piaget, 1971b) which nevertheless contains reports of several experiments.

The theoretical developments appear primarily in volumes of the *Études d'Épistémologie Génétique*. Unfortunately, little of this has appeared in translation, with the exception of Beth and Piaget (1961/1966) and the extracts selected by Furth (1969). However, some books are available in translation that describe many of these new developments, at least in summary form. Piaget and Inhelder (1967/1969) give a concise summary of the entire theory of intellectual development. Two books deal with the role of epistemology in Piagetian theory (Piaget, 1971a, 1970/1972). Piaget (1968/1971) discusses structuralist theories of knowledge (such as those of Levi-Strauss and Chomsky) and, in so doing, gives the clearest account available of the role the concept of structure plays in his own theory. In another book (1967/1971) he discusses the relationship of his theoretical approach with biology. It also contains an extensive discussion of the mechanisms he feels are responsible for cognitive development. These books provide a version of Piagetian theory which is sufficiently different from the empirically oriented view common in North America to warrant a detailed discussion.

2. A BROADER PERSPECTIVE

It is now necessary to come face to face with an important issue that we have allowed to slide by. It is impossible to discuss applications of Piagetian theory to educational practice meaningfully unless we state what we think we mean by Piagetian theory. Up to now, we have only done two things, neither of which is sufficient. The main phases of Piaget's career have been described, and a version of Piagetian theory prevalent in North America (the subset of the theory that can be obtained by taking those parts of Flavell's [1963] exposition that can be directly related to the experiments) has been characterized as inadequate. An attempt will now be made to suggest a more adequate version. In so doing, it will become clearer why the North American version is inadequate.

2.1 *Piagetian Theory and Scientific Method*

It will be necessary to begin by asking a preliminary question: Is Piagetian theory really a scientific theory at all? This immediately poses serious problems. It is clear from Piaget's recent writings that the theory is only in part a psychological theory. More generally, it is an all-encompassing theory of knowledge. "Genetic Epistemology attempts to explain knowledge, and in particular scientific knowledge, on the basis of its history, its sociogenesis, and especially the psychological origins of the notions and operations upon which it is based" (Piaget, 1971a, p. 1). As a theory of knowledge, it contains its own built-in notions of what science and scientific theories consist of. However, these notions are implicit rather than explicit and are intricately meshed with the psychological theory. "In sum, genetic epistemology deals with both the formation and the meaning of knowledge. We can formulate our problem in the following terms: by what means does the human mind go from a state of less sufficient knowledge to a state of higher knowledge? . . . The fundamental hypothesis of genetic epistemology is that there is a parallelism between the progress made in the logical and rational organization of knowledge and the corresponding formative psychological processes" (pp. 12-13). Broadly speaking, then, the growth of science parallels the growth of the mind. A Piagetian analysis of Piagetian theory might be intriguing to pursue, but it would not answer our question because of its obvious circularity. It may be true that Piagetian theory satisfies Piaget's criteria for what a scientific theory should be like. To assert that this implies that Piagetian theory is scientific would be using the theory to justify itself!

For a satisfactory answer to our question, it is necessary to turn to the history and philosophy of science. Unfortunately, Piagetian theory does

not fit into the classical framework of the philosophy of science, at least as this area is generally interpreted by psychologists. According to this viewpoint, a theory is a body of knowledge from which empirically testable hypotheses can be deduced. It exists to explain empirical facts, produced by well-controlled experimental procedures. If the facts clash with the predictions of the theory, the explanation is rejected. To fit into this framework, a theory must be clean. It must make sense as a deductive system and it must be possible to specify what parts of the theory must be abandoned or changed in the face of a given piece of negative evidence. Also, well-defined experimental procedures must exist with the capability of generating such negative evidence, should nature so decide. There are at least three problems with Piagetian theory in this regard. The first is that it is unclear whether the theory exists to explain empirical facts (except inasmuch as the existence of scientific knowledge is an empirical fact). It is more as if the facts exist as illustrations of the theory. The second is a lack of clear rules of inference. We are given many definitions, but few clues about how to make deductions from them. The third is a peculiar resistance to refutation. The literature, by now, abounds with "refutations" of aspects of Piagetian theory. A close examination invariably shows that it is most unclear what is being refuted. The difficulties involved in reproducing Piaget's clinical procedures have already been commented upon. There is, however, a more subtle issue. The attempt to generate sharp, testable hypotheses seems to lead to major distortions or oversimplifications in the theory itself.⁸

Fortunately, other frameworks are available. The most profitable, from our point of view, come from the history of science. In recent years, there has been a considerable amount of critical historical analysis of established scientific theories, especially Newtonian physics. It has become clear that most real theories (in their raw form as opposed to the form they take in textbooks) are not clean and do not fit into the classical framework either.⁹ In particular, they are never abandoned solely on the basis of negative empirical evidence.

The alternative framework best known to psychologists is that proposed by Kuhn (1962). Theories are collections of *paradigms*, or model

⁸ See, for example, Piaget's (1967) review of Bruner's research on cognitive development (Bruner, Olver, & Greenfield, 1966). An interesting account of how such distortions come about, by an author who is very much aware of what he is doing, is to be found in Osherson (1974).

⁹ Facile comparisons between Newton and Piaget should not be encouraged, however. The pathologies in Piagetian theory are on the surface, as it were, whereas the pathologies in Newton are buried beneath an elaborate deductive system and require elaborate historical analysis to be brought to the fore. Moreover, subsets of Newtonian theory do fit into the conventional framework. It is unclear, at present, whether the same is true for Piagetian theory.

ways of doing science that are accepted by some community of scientists.¹⁰ Such paradigms become accepted because they operate effectively in setting and solving problems of interest to that community. However, they are heuristics rather than algorithms and do not work successfully all the time. An instance of a paradigm not working when it should is called by Kuhn an *anomaly*. Negative empirical evidence no longer refutes a theory. Rather, it creates an anomaly for a paradigm. Paradigms are never correct or incorrect. Rather, they are adopted and dropped. Why should they be dropped? Because paradigms ultimately cease to be effective. In particular, they accumulate anomalies. At some point, there is a paradigm shift. A new paradigm is discovered and it supersedes the old one.

At this point, we can breathe a sigh of relief. Since all that is needed is a set of heuristics accepted by a community of scientists, Piagetian theory is clearly scientific by Kuhn's standards (as long as we are willing to acknowledge that psychologists and educational researchers are scientists). There are other advantages to viewing the theory paradigmatically. For example, the reasons for Piaget's rediscovery become clearer if we view it as part of a major paradigm shift that occurred in North American psychology between 1955 and 1965. Also, the theory loses some of its opaqueness. For example, the notion of developmental stages can be clarified by viewing it as a combination of two paradigms: (a) a way of generating empirical demonstrations from epistemological considerations and (b) a way of classifying age-related differences in children's performance in these demonstrations.

However, we are not yet out of the woods. As Lakatos (1972) has pointed out, there are flaws in Kuhn's framework that are especially serious for theories that, like Piaget's, persist in the face of accumulated negative evidence or anomalies. As both Kuhn and Lakatos point out, such theories often have a remarkable longevity (the most glaring example being Ptolemy's geocentric astronomy). In Kuhn's analysis, paradigms are dropped because of a feeling of dissatisfaction about their productivity rather than the existence of anomalies per se. Nonetheless, the possibility must be faced that a theory with serious anomalies is a bad theory. This is particularly important because Kuhn's definition is so broad that it admits the possibility of any theory, however bizarre, as long as it is accepted by some community of scientists. The only safeguards are

¹⁰ A paradigm is being viewed here as a *procedure*. A procedure can be viewed as consisting of subprocedures, so a paradigm can be viewed as consisting of subparadigms and a collection of subparadigms defines a paradigm. Hence, the terms "theory" and "paradigm" are interchangeable. There has been considerable controversy over what Kuhn himself actually means by "paradigm." This definition is intended to capture the spirit rather than the letter of Kuhn (1962). It may not be entirely consistent with Kuhn's more explicit definitions in the postscript added in 1969 to the second edition of his book.

sociological. Is there a consensus for it or against it? It would be meaningless to evaluate Piagetian theory according to this criterion. The result would vary drastically depending on the group of investigators being talked to and the aspect of the theory being talked about. If, on the other hand, the criterion is abandoned, then Kuhn's definition of a scientific theory is not really a definition at all—it admits any way of doing anything. Rather than resolve this dilemma, it is better to look for a more convenient framework while retaining the terminology that Kuhn introduced.

We will adopt the framework proposed by Lakatos (1972). According to this point of view, there is no such thing as a single monolithic theory. What is usually referred to as a theory is really a sequence of theories that change over time. Lakatos calls such a sequence a *research program*. A research program has two components: an invariant *hard core* and a changing *protective covering*. Thus, the component theories differ only in the content of their protective covering. They contain the same hard core, and it is this hard core that characterizes the overall research program.

Like Kuhn, Lakatos views science as a problem-solving activity. However, the content of science is more highly structured. Elaborating slightly on Lakatos, we may distinguish between three types of content: heuristic procedures (essentially Kuhnian paradigms but without the element of consensus), hypotheses, and empirical data. A research program may be viewed as a vast procedure for generating hypotheses. The hypotheses so generated are called *auxiliary hypotheses* and constitute the constantly changing protective covering. They are generated by two kinds of heuristics: positive and negative. The negative heuristics are designed to protect the invariant hypotheses in the hard core by rendering them incapable of refutation. They respond to anomalies by changing auxiliary hypotheses in the protective covering.¹¹ The positive heuristics are there to give coherence to the research program. They form a partially articulated plan for generating new auxiliary hypotheses and thus elaborating the protective covering, more or less independently of any anomalies that may arise.

This framework is closer to the classical point of view. The hard-core hypotheses are like axioms, the heuristics are like rules of inference, and the auxiliary hypotheses are like empirically testable propositions. However, there are three major differences. The first is the view that the proper unit of analysis is a sequence of theories rather than a single theory. The second is the greater prominence given to procedural notions. The third is that empirical data per se do not play a major role in the development of a research program. In fact, Lakatos gives a number

¹¹ Such a change may take a number of forms. It may be a major modification of the theory, it may simply be an ad hoc way of rendering the anomaly harmless, or it may constitute an attack on a rival research program.

of examples of research programs, whose hard-core hypotheses are nowadays accepted as valid, which survived for many years "awash in an ocean of anomalies." The positive heuristics are more critical than the negative ones. To Lakatos, a research program is good to the extent that changes in auxiliary hypotheses result in extensions of the range of empirical phenomena that can be accounted for, but only occasional empirical corroboration is necessary. The important thing about the protective covering is not the testability of the auxiliary hypotheses but the fact that they directly say things about empirical phenomena. In the long run, the protective covering protects because it displays the power of the hard core.

Lakatos's framework is not without its flaws.¹² However, it provides a tool for explaining how a perfectly legitimate scientific theory can handle anomalies yet remain viable. The major conceptual problem with Piagetian theory is to explain precisely this fact. It will be seen that it is also valuable in other ways. The division into hard core and protective covering can be used to clarify which aspects of Piagetian theory are of critical importance and which are not. It also provides a critical apparatus for examining the relationship of Piagetian theory to other research programs (when confusion does not arise, we will use the terms "research program" and "theory" interchangeably). What we have referred to as the North American version of Piagetian theory is an example. If we form a theory out of the empirically oriented aspects of Piaget's work, we have a new research program. It may have the same protective covering, initially, but it will not have the same hard core. In fact, it may only have a trivial hard core. Most important, it is extremely unlikely to have the same negative heuristics. Obviously, such a research program is likely to evolve in a different direction. In any case, the inability of such a program to cope with an anomaly does not necessarily mean that Piagetian theory is unable to cope with it.

2.2 *The Hard Core of Piagetian Theory*

We are now in a position to state what we think we mean by Piagetian theory. It is a research program with a hard core, positive and negative heuristics, and an ever-changing protective covering. In terms of this framework, it is unnecessary to describe the protective covering. It is only necessary to describe the heuristics and the hard-core hypotheses in sufficient detail to indicate how auxiliary hypotheses can be generated. The present description should be viewed as only a rough first approximation to a proper formulation along these lines. It is schematic and incomplete. It is not intended as a basis for a new research program, nor

¹² See Laudan (1977) for an analysis of some of these flaws, and also some of Kuhn's.

is it intended to serve as a substitute for Piaget's own description of his theory. It is intended for one purpose and one purpose only, namely, to define as concisely as possible what will be meant by "Piagetian Theory" for the remainder of this paper.

We will adopt the position that the hard core is to be found primarily in the epistemology. I will state the hypotheses in my own words rather than Piaget's. Most of them are paraphrases of statements to be found in *Structuralism* (1968/1971), *Genetic Epistemology* (1971a), or *Biology and Knowledge* (1967/1971). Before turning to the hypotheses themselves, it is necessary to consider the meaning of three terms that are basic to Piaget's epistemology: state, transformation, and structure. Piaget's use of *state* seems identical to its use in science and mathematics as a term for a condition that something is in or a property that something has. A *transformation* is a law or procedure that, when applied to a state, determines or generates a new state.¹³ Whereas a state is to be viewed as something static, a transformation is to be viewed as something dynamic (in much the same way as a computer program is dynamic, whereas the data used as input are static). It is definitely something that works or functions rather than merely exists. A *structure* is a system consisting of a set of states (which a mathematician would call the state space), a set of transformations, and a set of global laws governing the application of the transformations. It has the following properties:

(a) The system is a "totality."¹⁴ The global laws apply to all elements of the system.

(b) The system is "self-regulating." The state space contains every state which a transformation can act upon. The functioning of the transformations is completely determined by the global laws within the system.¹⁵

(c) Structures are modular. A given structure can contain substructures, each with its own state space, transformations, and global laws. Conversely, if a structure is a substructure of a larger system, its boundaries remain intact and its laws are conserved.

¹³ The mathematically inclined reader might think that this is the same as a function or partial function. However, there is a subtle but important distinction between the mathematical notion of a function and Piaget's notion of a transformation, analogous to the distinction between a function and an effective procedure for computing its value.

¹⁴ Terms in quotes are Piaget's own or those invented by his translators.

¹⁵ It is important to note that the state space may be only implicitly defined, for example, as the set of all possible outputs of the transformations. Thus, we *may* only have a general characterization of the state space and the transformations may generate "new" states that have never been encountered before. It should also be noted that self-regulation, defined this way, does not necessarily imply closure (i.e., transformations could conceivably generate states outside the state space). Piaget frequently states that structures are closed (e.g., 1971a, p. 14). However, this is usually in the context of a discussion of mature or highly organized knowledge. He also seems to think that transitional structures (e.g., the "pre-structures" of pre-operational children) exist with only partial closure. Here, the same terminology is used for both kinds of structure.

HYPOTHESIS 1. Reality, by which Piaget seems to mean physical phenomena (including the overt behavior of living organisms), consists of objects, states of objects, and transformations of states. Any phenomenon that persists or recurs corresponds to a set of transformations that leave some states unchanged.

HYPOTHESIS 2. All biological and psychological processes are structures. Such structures can be said to *function* whenever a transformation is carried out or executed.

HYPOTHESIS 3. Knowledge consists of structures. In particular, knowledge of reality consists of structures that form "models" of reality.¹⁶ Scientific knowledge is of this type. On the other hand, mathematical knowledge (logico-mathematical knowledge in Piaget's terminology) consists of knowledge about arbitrary structures. In other words, it consists of structures that form models of other structures.¹⁷

HYPOTHESIS 4. Some structures are better than others in their potential adequacy to form models of reality. The best are those in which the transformations are *operations*. An operation is a transformation that is *reversible*. Reversibility is difficult to define. It is not quite the same as the notion of a transformation having an inverse. The gist is that an operation is reversible if it can be put into reverse and so reconstruct the states that it originally transformed, without undoing the effects of the original operation.

HYPOTHESIS 5. Organized rational knowledge (science, mathematics, etc.) and the processes involved in thinking obey parallel laws. In particular, they obey parallel laws of development.

HYPOTHESIS 6. A structure (taken in isolation) can develop in two ways: (a) The range of phenomena or other structures that it models can increase (assimilation), and (b) changes can occur in the structure itself (accommodation). Assimilation and accommodation are complementary. One cannot occur without the other.

¹⁶ "The transformational structures of which knowledge consists are not copies of the transformations in reality; they are simply possible isomorphic models among which experience can enable us to choose" (Piaget, 1971a, p. 15).

¹⁷ It is difficult to pin down exactly what Piaget means by a model. Since the notion is very important, I tentatively offer the following definition: (a) A model of reality is a structure which is a model for a physical phenomenon; (b) a structure X is a model for a phenomenon P if every transformation in P corresponds to a transformation in X ; (c) structure X is a model for structure A if there exists a mapping (of states and transformations) from A to some substructure S of X such that S has the same global laws as A . Category theory (e.g., MacLane, 1972; Arbib & Mannes, 1975) is an example of a branch of mathematics that deals explicitly with structures that represent other structures by means of mappings of this kind.

HYPOTHESIS 7. The structures of thought maintain themselves,¹⁸ and knowledge develops because of these efforts at self-maintenance. If a structure can model all phenomena it encounters, it can be said to be in equilibrium. When equilibrium is destroyed or threatened, the individual must attempt to regain it. This can be done in two ways: (a) by developing the structure itself according to Hypothesis 6 or (b) by building or extending new structures that coordinate the given structure with others. These latter are the logico-mathematical structures of Hypothesis 3. The process by which they are formed is called by Piaget in his more recent writings *reflective abstraction*.

HYPOTHESIS 8. Operational structures (i.e., those in which all the transformations are operations) are the only ones that can achieve permanent equilibrium. In particular, they allow for errors to be anticipated and corrected before they are made. Other structures do not have this capability.

HYPOTHESIS 9. Many structures survive without achieving permanent equilibrium.¹⁹ They persist unchanged because there exist operational structures that coordinate them.

HYPOTHESIS 10. At any point in time, the knowledge that a child can acquire immediately is limited by the structures that are available to him. All the infant has available at birth are some preprogrammed structures, the capacity to perform actions, and the capacity to form structures (the "sensory-motor schemes") that are internal representations of states, actions, and the coordination of actions.

HYPOTHESIS 11. Knowledge develops because the sensory-motor schemes develop and coordinations between schemes develop. These coordinations are themselves structures, which can be coordinated once again. Thus, the pattern of development is hierarchical, consisting of the formation of structures of structures. Every time a new level of the hierarchy is established, a major reorganization of thought has taken place.

¹⁸ "The operations of thought, taken as a whole, can be considered as a vast autoregulatory system which gives autonomy and coherence to the process of thought" (Piaget, 1967/1971, p. 11).

¹⁹ Those involved in the "figurative aspects" of thinking (as opposed to the "operative aspects") are of this type. "The figurative aspect is an imitation of states taken as momentary and static. In the cognitive area the figurative functions are, above all, perception, imitation and mental imagery. . . . The figurative aspects are always related to the operative aspects. Any state can be understood only as the result of certain transformations or as the point of departure for other transformations" (Piaget, 1971a, p. 14). These figurative aspects might be viewed as being analogous to methods of collecting and analyzing data and the role they play in scientific knowledge.

HEURISTIC 1 (The Epistemological Base). Find correspondences between mathematical and scientific knowledge, on the one hand, and children's thinking, on the other. Analyze the areas of knowledge so isolated to find what structures appear necessary and sufficient for the knowledge to exist.

HEURISTIC 2 (The Importance of Groups). Mathematical structures in which the transformations have inverses correspond to mental processes that can achieve equilibrium.

HEURISTIC 3 (The Formal Theory). As an aid in developing these correspondences, develop systems of notation²⁰ that indicate what transformations are involved and what is the inverse of what.

HEURISTIC 4 (Stage Theory). Use the correspondences obtained by Heuristic 1 to identify tasks that require the use of a given structure. Classify tasks into stages according to the position of the structure in the hierarchy of Hypothesis 11. Classify potential errors that might be made on such tasks by identifying structures that would cause such errors to be made.

HEURISTIC 5 (The Clinical Method). Classify a child's performance on a given task on the basis of (a) the classification of task and potential errors obtained from Heuristic 4 and (b) his responses to probes designed to ascertain the underlying structures responsible for his performance.

NEGATIVE HEURISTIC 1 (Epistemological Triviality). Experiments involving a task that does not have an explicit epistemological base (in the sense of Heuristic 1) are irrelevant to the research program.

NEGATIVE HEURISTIC 2 (D  calages). The hard core defines a "main sequence" of development. Structures may arise outside this main sequence, but they are either transient or survive only by being incorporated into structures in the main sequence. Look for such structures to explain empirical results that are inconsistent with the predictions of Heuristic 4.

Ideally, these hypotheses and heuristics should be sufficient to generate the research program. Everything else usually ascribed to the theory has the status of an auxiliary hypothesis. It should be noted that there are a number of different kinds of auxiliary hypotheses. There will be epistemological hypotheses and hypotheses dealing with developmental mechanisms as well as purely empirical hypotheses. The reader with an

²⁰ Examples are the *groupements* and the INRC group. It is important to emphasize that if this view is correct, and they actually are languages designed to be of heuristic value in establishing correspondences between thought and organized knowledge, then it is meaningless to view them as mathematical models, in the sense in which psychologists usually use the term.

interest in the history and philosophy of science should also note that a deliberate compromise has been made between the need to give a self-contained but comprehensible statement of the theory and the need for the hard-core hypotheses to be invariant over time. These may not be the hypotheses that Piaget started out with at the beginning of his second phase. However, it can be argued that he started out with something very much like them.

3. SOME IMPLICATIONS FOR EDUCATION

We now come to the second main section of this essay. We have considered Piaget's career, have given some general context about how his work came to be so widely recognized, and have attempted to provide an explicit characterization of the theory. We have seen that Piaget's recognition in North America was accompanied by a great deal of anticipation regarding its potential applications. We are now ready to consider what this potential consisted of.

In so doing, it will be useful to distinguish between movements and specific applications. It is clear that Piaget has had an enormous influence on two movements: the curriculum reform movement in the United States and the movement toward child-centered primary schools in England. The origins of the first have already been discussed. The second owes its origins to the role of N. Isaacs²¹ and the National Froebel Foundation in the early dissemination of Piaget's work. However, movements are diffuse and difficult to characterize, and where or when a particular idea occurs is difficult to document. While an adequate analysis of Piaget's influence on these movements is certainly of vital importance, it will not be attempted here. It would take an entire book to do justice to either.

When we turn to specific applications, we are faced with a different problem. They are apt to involve different interpretations of Piagetian theory, and these interpretations frequently bear little relationship to each other, at least on the surface. It is as if Piaget means all things to all men. Such interpretations are frequently confused or erroneous. There is a tendency to confuse the auxiliary hypotheses with the hard core of the theory. There is also a tendency to select a few portions of the theory and combine them with portions of other theories that are completely different, and in some cases even incompatible. It is therefore difficult to survey the applications in anything but a piecemeal fashion. It would also be difficult, in such a survey, to avoid extensive discussions of a wide variety of other theories and their educational implications, without doing

²¹ Two of Isaacs' most influential papers are reprinted in Isaacs (1974). Many of the Piagetian ideas that have been most influential are discussed by Elkind (1976) in a slightly different context.

an injustice to the applications themselves.

Fortunately, an alternative approach is possible. This is to examine what Piaget himself has written about education and to consider the relationship of these ideas to his overall research program, the practical techniques that they suggest, and the difficulties encountered in implementing these techniques. It will be seen that a surprisingly concrete picture emerges of the *potential* contributions of Piagetian theory to education. It may also become somewhat clearer why this potential has proved so difficult to actualize.

3.1 *Piaget on Education*

Piaget has not written any books on education per se. What he has done, over the years, is to write a series of essays. These are of two kinds. Some deal in a statesmanlike manner with very general policy issues only indirectly related to the theory. They need not concern us here. The others deal more specifically with the relationship of the theory to practical educational matters that might be encountered in schools. However, the treatment is always at a quite general level, and specific applications are almost invariably only mentioned in passing. Three of these essays appear in translation in two books (Piaget, 1971b, 1974).

The earliest is *The New Methods: Their Psychological Foundations*, which was written in 1935. It appears in Piaget (1971b). The new methods are those that were used in the progressive education of the 1930s, which emphasized "learning by doing" or fostering activity in the child. Piaget is in favor of progressive education. In the traditional school, the child's activity is inseparable from a continual constraint exercised by the teacher. The progressive school, on the contrary, appeals to real activity, to spontaneous work based on need and interest. However, there is a problem. In the traditional system, a child's behavior is to a great extent predictable and controllable by the system itself. If, on the other hand, use is made of "progressive" methods where a child is free to adapt to his environment, then a much more precise knowledge of the child is needed in order to understand how and when they are effective. In particular, careful analysis is needed of four issues: the nature of childhood, the structure of the child's thought, the laws of development, and the mechanism of infantile social life.

The rest of this early essay is devoted to an account of how Piagetian theory can shed light on these four issues. The nature of the child is defined by what we have called the hard-core hypotheses of the research program. The structure of thought is the epistemology in an early form. The laws of development refer, of course, to stage theory. The mechanisms of social life involve the egocentric thought that Piaget (1932) feels is characteristic of the young child.

Taken as a whole, it is an exposition of principles that might be applicable. It is not a description of how to apply them. However, there are two very important hints. The first is the notion that play is crucial in ensuring the assimilation of reality.²² The second is a warning against too great a neglect of the environment, especially its more structured aspects.²³ Thus are introduced two themes that play a crucial, and complementary, role in the success or failure of any attempt to apply Piagetian theory. Free activity is important, and so is the structure of the environment. Yet the more explicit a structure we put on the environment, the less opportunity a child has for free activity and vice versa. The critical issue in any Piagetian application is: How does one achieve a satisfactory balance between these two apparent opposites?

The second essay is *Education and Teaching Since 1935*. It, too, appears in Piaget (1971b) but was written in 1965. Intended as a sequel to the first essay, it is addressed to the question of why there exists no "science of education sufficiently developed to be able to answer the innumerable questions that present themselves every day and whose solution is therefore left either to empiricism or to tradition." To Piaget, this is a policy issue as well as a scientific one. The situation he would like to see would be analogous to that in the field of medicine. Ministers of health do not legislate in the sphere of medical knowledge because there already exists a science of medicine, with the authority to supply both principles and factual data, leaving to the ministry only those problems concerned with determining the best applications. On the other hand, the ministry of education decides not only upon applications but also upon the principles to be applied.²⁴

Much of Piaget's analysis in this essay deals with administrative and political problems. However, he does devote a considerable amount of attention to one issue that is closely related to direct applications of the theory. This is the tendency for educational research to exist in a state of isolation from scientific disciplines that could support it with the principles it needs. Thus, educational research may have quite powerful and

²² "For assimilation in its purest form . . . is in effect nothing other than play; and play, which is one of the most characteristic infantile activities, has in fact found a use in the new techniques of education for young children that remains inexplicable unless one clarifies this function to the child's mental life as a whole and to its intellectual adaptation" (Piaget, 1971b, p. 155).

²³ "Suppose . . . that the development of individual thought is comparable to an embryology obeying strict hereditary rules, then the consequences would be incalculable for education: the teacher would be wasting his time and his effort attempting to speed up the development of his students, and the problem would simply be that of finding out what knowledge corresponded to at each stage and then to present it in a manner assimilable by the mental structure of the age level in question" (Piaget, 1971b, p. 167).

²⁴ Here, Piaget is referring specifically to the situation in Switzerland and France. However, it seems reasonable to assert that the same situation exists in the United States. The best example is the tangled relationship that has existed between Congress and the National Institute of Education over the last few years.

sophisticated evaluation techniques available, but the ability to explain the results is almost totally lacking. To Piaget, only an interdisciplinary approach can provide such explanations. In particular, it is obvious, to Piaget, that educational research will have to employ a precise psychology, not merely that of common sense (1971b, p. 24).

Piaget proceeds to outline how his own approach can contribute to a precise psychology of this kind. He begins by restating the basic problem of the 1935 essay in a more mature form. There are "traditional" methods of education, based on the notion of gradually incorporating children into adult social life and determined by adult norms. There are "progressive" methods, which emphasize the spontaneous and relatively autonomous aspects of the child's intellectual structures. Presumably, these are determined more by what a child can actually do well. However, the distinction is no longer a pejorative one. To Piaget, the traditional methods are perfectly good for teaching things such as spelling that do not have any real epistemological base but are simply things that are arbitrarily defined by the norms of adult society. The issue is whether there might not be better methods for teaching mathematics and science²⁵ and other areas that have developed through a process of research and discovery.

To Piaget, such an area can be taught in two ways: passively, as an assemblage of facts and rules, or actively, by means of some process that involves the reconstruction or reinvention of facts and rules. He immediately makes it clear that he prefers the second way on ideological grounds. The question he is interested in dealing with, however, is whether there exist good underlying psychological reasons for this preference.

Piaget proceeds to indicate how his research program provides some such reasons. The argument is complex. Its gist is as follows. The principal aim of education is to develop intelligence. The essential functions of intelligence consist of understanding and inventing. In terms of the hard core of the research program, understanding a phenomenon consists of generating transformations belonging to a structure that models it. Invention consists of developing the structure. In much the same way (it would seem) as assimilation and accommodation are complementary, so are understanding and invention, and intelligence itself is a condition of equilibrium that results from their functioning. Empirically, the development of intelligence corresponds to the sequence of stages provided by stage theory. Theoretically, it corresponds to the development of structures with an ever-increasing capacity to assimilate and organize (or model) reality. This development depends on natural, or spontaneous,

²⁵ "A mathematical truth is not dependent upon the contingencies of adult society but upon a rational construction accessible to any healthy intelligence; an elementary truth in physics is verifiable by an experimental process that is similarly not dependent upon collective opinions but upon a rational approach, both inductive and deductive, equally accessible to that same healthy intelligence" (Piaget, 1971b, p. 26).

processes whose existence constitutes a necessary condition for the effectiveness of any form of instruction. They may be utilized and accelerated by education, but they are not derived from it. However, Piaget once again takes pains to warn us that the fact that these processes are natural and spontaneous does not imply that they are determined solely by maturation. In fact, little is known about the role of maturation, according to Piaget, except that it exists. It can do no more than open up possibilities. The theory is about how such possibilities become actualized.

This brings us to the important point. It follows from Piagetian theory²⁶ that these natural and spontaneous processes are analogous to the methods of research used in developing science and mathematics. Indeed, the child can be viewed as carrying out his own research program and using it to construct his own sequence of theories about reality.²⁷ That is how he or she assimilates knowledge. There are, of course, other things involved in the acquisition of knowledge. In particular, knowledge can be communicated through books, speech, pictures, symbols, or demonstrations. Hence, traditional methods have their place. However, structures must be developed to assimilate or model this knowledge and this can only be done through mental or physical activity involving these natural and spontaneous processes. Hence, the curriculum should make room for such activities. Most important, the opportunity for these activities must be extensively provided in science and mathematics because it is only through the performance of such activities that these subjects can be properly understood.

3.2 *Piagetian Techniques*

Up to now, we have been concerned more with what Piaget has to say about educational principles than with what he has to say about educational practice. The rest of the second essay (or rather that portion not concerned with educational policy) deals with the implications of these principles for practical educational techniques. The short third essay, entitled *A Structural Foundation for Tomorrow's Education* (Piaget, 1974), is entirely devoted to this issue. There is also a book which was not written by Piaget himself but by a student of his (who happened to be an experienced school teacher) for a doctoral thesis (Aebli, 1951). It is important because it deals specifically with education, from a rather practical point

²⁶ See especially Hypothesis 5 and Heuristic 1 in the section of this paper that deals with the hard core.

²⁷ Lakatos's viewpoint is not inconsistent with Piaget's. The hard core gives us a structure, in Piaget's sense of the term. The hypotheses are states and the heuristics are transformations that, when executed, generate auxiliary hypotheses. The notion of an invariant hard core is not unlike some of Piaget's notions of invariance. At the risk of pushing things too far, it might even be said that the notion of a sequence of theories that progress in their ability to generate new empirical facts resembles some of Piaget's notions about how structures develop.

of view, and is the first treatment (at least by anybody associated with Piaget) of what later came to be called "discovery learning." Many of the ideas that later appear in Piaget's essays first appeared in this book. Curiously enough, it is almost unknown in North America.²⁸

We will not discuss these sources separately. If put together, it is possible to abstract out two techniques. Although they are quite general,²⁹ their specific implementation by others constitutes (together with applications of stage theory, which will be treated in a separate section) the kernel of most Piagetian applications. One is concerned with how to teach. The other is concerned with what to teach. A specific application may make use of only one of these techniques (usually the first), but both are important. They constitute the two aspects which, as was pointed out earlier, it seems necessary to balance in any successful application. The first defines the desirable free activity. The second defines the desirable structure. Ignoring the second can lead to aimless play. Ignoring the first can lead to a sophisticated curriculum that most students fail to assimilate or understand. While the two will be discussed separately, it should always be borne in mind that each implies the other. At the end of this section, we shall attempt to bring the two together.

The first technique consists of putting a child in what might be called a free problem-solving situation. The child selects a task and is free to do anything he likes in order to accomplish it. The teacher can (and at times must) intervene, but it is absolutely critical that the child's activities remain spontaneous. There are two models for how this intervention should occur. The first is the Socratic dialogue. The teacher can guide by asking appropriate questions. The second is that of the child as an experimenter³⁰ and the teacher as a research director. The whole purpose of the technique is to put the child into a situation where he or she can invent hypotheses and put them to the test. There is a very crucial analogy here between what a child does and what the scientist or mathematician does. A scientist changes his mind only if he is aware of an anomaly. A mathematician changes his mind only if he is aware of a counterexample.

²⁸ It remains untranslated. To my knowledge, it is not cited in any of the standard books on discovery learning (e.g., Shulman & Keislar, 1966). There is, however, a brief summary in Flavell (1963).

²⁹ They might more accurately be called heuristics or paradigms (in the sense of desirable ways of doing education).

³⁰ It is absolutely critical to note that these are not necessarily experiments with physical objects. They may be conceptual experiments, or the testing of mathematical hunches by means of pencil and paper. To quote Piaget (1974) on this point: "An active school is not necessarily a school of manual labour, and, although the child's activity at certain levels necessarily entails the manipulation of objects . . . at other levels the most authentic research activity may take place in the spheres of reflection, of the most advanced abstraction and of verbal manipulation (provided they are spontaneous and not imposed on the child at the risk of remaining partly uncomprehended)" (p. 68).

A child will learn only if he extends the range of hypotheses he can generate and modifies or eliminates the transformations that lead to false ones. Thus, it is part of the teacher's task to ensure that the child is aware of anomalies and counterexamples that result from his activities.

The second technique is what might be called constructive analysis. A class of tasks and activities is selected and their cognitive demands are analyzed. The goals are to provide some kind of natural ordering of their complexity to specify the kinds of structures a child would have to possess in order to benefit from them and to describe how they can facilitate the construction of new structures. There are at least three kinds of analysis that can be used. They are not unrelated. One is Piaget's own method of classifying complexity through the use of Heuristics 1 and 4 of the hard core of the theory. Another is to develop parallels with the history of science. The third is to make use of certain notions that arise in modern abstract mathematics and are used there to order the complexity of mathematical concepts and structures.³¹

The descriptions of both techniques are deliberately idealized. They are extremely difficult to apply in practice. Indeed, without a considerable amount of research on how to use them and how to train others in their use, they may be directly usable only by mathematically sophisticated research scientists who also happen to be excellent teachers. It is not surprising, therefore, that most actual applications fall far short of this ideal. In the case of active problem solving, the shortcomings have tended to be in the direction of supplying inadequate guidance and failing to point out counterexamples and anomalies. In the case of constructive analysis, the tendency has been to ignore any original epistemological analysis and opt for simpleminded classifications of complexity based on the auxiliary hypothesis of stage theory.

These shortcomings are, to some extent, unavoidable and are not necessarily bad. There is a reason, however, for stressing the idealized versions of the techniques. It seems reasonable to assert that they specify the necessary (though not sufficient) conditions for rationally ensuring the balance between freedom and constraint that was mentioned earlier in this section. As the shortcomings increase, balance becomes more and more a matter of luck. It becomes harder to tell whether the most appropriate tasks are being used. It becomes harder to specify exactly how a child profited from any given activity.

There is one kind of activity where the application of these techniques is somewhat easier. This is the kind that is naturally self-correcting. The counterexamples and anomalies are generated by the activity itself. This

³¹ One example is category theory (MacLane, 1972). Another is the approach developed by the Bourbaki school of French mathematicians. This latter approach has been particularly influential on Piaget's thought. See the essay *Mother Structures and the Notion of Number* in Ripple and Rockcastle (1964).

removes one of the most difficult aspects of Piagetian teaching: the need for the teacher to point out that there is something wrong with a child's hypotheses by means that do not interfere with the child's spontaneous activity. It is also frequently easier to develop intuitive notions of task complexity. The most obvious activities of this kind are physical ones (and this may be the reason why so many Piagetian applications put such a heavy emphasis upon physical activity). For example, the child learning to ride a bicycle knows when he has made an inappropriate action because of the way the bicycle behaves and what this does to the relationship with the environment (an anomaly may literally hit him in the face if it happens to be a lamppost). The point is not the existence of feedback per se or the fact that it is immediate. The point is that the nature of the anomaly varies with the nature of the error. Such activities do not necessarily involve the mastering of a physical skill. Certain games have this property. So has computer programming. We will return to a consideration of this latter activity in our discussion of Seymour Papert's LOGO environment.

3.3 *Stage Theory*

There is one other technique that can characterize Piagetian applications. This is the direct use of stage theory. In our discussion of the hard core of Piagetian theory, it was viewed as a heuristic. It can, of course, also be viewed as a very rich set of auxiliary hypotheses about what children of various ages or developmental levels tend to be able to do. In a sense, these auxiliary hypotheses have taken on a life of their own. They can be accepted as empirical facts and embedded in theoretical approaches quite different from Piaget's own. This is why this technique is being considered separately from the other two, which are much more closely linked to the hard core of Piaget's own research program.

Basically, the technique consists of (a) estimating the child's developmental level and (b) deciding what constraints this imposes and what to do about these constraints. Two issues will be considered in this section. The first is how developmental level can be practically measured in educational settings. The second is its stability. We will finish with a brief discussion of Kohlberg's (1975) theory as an illustration of the heuristic-use of stage theory in a research program different from Piaget's own.

The most obvious way of estimating developmental level is by inferring it from a child's age. This is dangerous, however, because it is well established³² that there are quite wide cultural differences in the average age at which a given stage is attained. There are also, of course, individual differences within cultures. Moreover, a child may be at a given stage only with respect to certain tasks. Hence, it is necessary to devise testing

³² See, for example, the review in Flavell (1963). An interesting discussion of the relationship between Piagetian notions of intelligence and more conventional notions, which pays particular attention to cultural difference, appears in Ginsburg (1972).

procedures.

Since the developmental levels of stage theory are interpreted within the Piagetian framework as levels of intelligence, such procedures might be viewed as being similar in intent to such standard intelligence tests as the Stanford-Binet (Terman, 1916). Inhelder (1943/1968) has described what a Piagetian approach should consist of. She objects to the Stanford-Binet and similar tests primarily on the grounds that it is unclear what is being measured and that the mental-age scale does not express a natural sequence of development (that is, it is unclear how a mental age of 10 evolves from a mental age of 9). A Piagetian test should be superior because it is specifically related to a theory of development. Thus, what is needed, according to Inhelder, is a testing instrument that (a) will analyze the psychological processes themselves, not just the output and the performances that result from them; (b) is embedded in a specific theory; and (c) yields data that are susceptible to a developmental interpretation.

Currently, a number of such testing instruments exist, or are close to existence. The best known is the Hunt-Uzgris scale of infant development. The scale for school-age children devised by Pinard and Larendau (1964) is closest to the Geneva approach in its use of Piaget's clinical method. The most ambitious is the British Intelligence Scale (Elliott, 1976). Such tests are criterion referenced rather than norm referenced. They involve, in one way or another, an adaptation of Piaget's clinical method to standardized procedures. The end result is the assignment of an individual to developmental stages and substages.³³

The use of such tests is not without dangers. As we have commented earlier, the auxiliary hypotheses of stage theory are, to some extent, task specific, and a child may be in different stages with respect to, say, different conservation tasks. The teacher who thinks that a child is deficient in a particular skill because the child is at the wrong stage may find information from even the best standardized test misleading. Of course, the teacher may apply his own test, and a way to do this is provided by Fogelman (1970). This is a compilation of the more important Piagetian tasks. Each is given in just enough detail so that the teacher can use one as a test, and there are just enough normative statistical data provided to render the results interpretable. However, it has a fatal weakness. The problem is that, while the tasks are described, the probes are not. Moreover, no mention at all is made of the existence of *décalages*, and the teacher is not cautioned to watch out for them and follow up with

³³ The simplest way to do this is by qualitative assignment, as is apparently done by Pinard and Larendau. A more quantitative approach results in severe problems, and traditional psychometric techniques are inappropriate. Some interesting problems arising from correlational analysis of test data are discussed by Tuddenham (1971). Other problems are pointed out by Elliott (1976). One is a high chance-success rate. Another is how to take care of *décalages*. Such problems may be surmountable through the use of Rasch scaling techniques (e.g., Rasch, 1966), which is the approach being used in the British Intelligence Scale.

more extensive testing. The use of raw Piagetian tasks, by a teacher unfamiliar with Piaget's clinical method, could yield an impression about a child that was far more misleading than the impression conveyed by a set of summary scores on some standardized test.

Having discussed the problems involved in determining what stage a child is in, we can turn to the problem of what to do about it. This is closely linked with the issue of the stability of stages. There are two extreme views on this matter, neither of which is Piaget's. The first assumes that nothing can be done and that all we can do is wait until the new stage appears. The second view, most vociferously held by Engelmann,³⁴ is that it simply reflects the operation of a learning process, and that the stage a child is in can be manipulated by appropriate training techniques. Either of these approaches leads to a complete lack of the balance we have mentioned as being so important in Piagetian applications to education. The first results in completely free activity. The second results in complete constraint.

Piaget's own position lies between these two extremes. Intervention of the right kind may have an effect, but it could take a very long time and should not interfere with the child's spontaneous activities. It is training, not learning, that Piaget is skeptical about. The only kind of training that would be likely to have a nontrivial effect on major transitions between stages would be the free problem-solving activity described in the preceding section. Some of its effects upon learning in class inclusion and conservation tasks have recently been examined by Inhelder, Sinclair, and Bovet (1974).

From the point of view of Piagetian theory, most training studies have been barking up the wrong tree.³⁵ The issue is not whether training procedures exist that improve performance on Piagetian tasks. The issue is what transitional structures are formed as a result of the training procedure and whether they can form a basis for further structures to develop. The nature of such structures and their effect on development form one of the major unexplored areas of the Piagetian research program. However, this does not mean that the problem is undefined. Major stage transitions correspond to major structural reorganizations. The problem is to explain how and why such reorganizations occur.³⁶

What follows from all this is that it is the structures that are of primary

³⁴ Engelmann's position is simply that the auxiliary hypotheses of stage theory are interesting facts that can be explained by a completely different theory. An interesting example of a clash between two rival research programs is to be found in the analyses by Engelmann (1971) and Kamii and Derman (1971) of the same training study.

³⁵ That is why they are essentially being ignored in this paper. The interested reader is referred to the review articles by Beilin (1971a, 1971b).

³⁶ "Stages are characterized by overall structures which become necessary but are not so initially. Formal structures become necessary when the concrete structures are complete; concrete structures become necessary when the structures of identity, of functions etc., are

importance, rather than the stages and the tasks used to define them. A test result indicating that a child is "in" a stage only has significance in that it indicates the minimal structures available. The fact that some training procedure improves performance says little, of and by itself, unless an attempt is made to discover what has happened from a structural point of view. Moreover, training studies may have little practical educational significance if they focus exclusively on the auxiliary hypotheses of stage theory (in much the same way as a project designed to increase performance on a subtest of the Stanford-Binet might lack educational significance). The stages were designed to characterize intelligence rather than knowledge. The acquisition of knowledge involves other cognitive functions, and their development follows different patterns of stages (Piaget, 1971b). It is the interaction of all these functions, and not intelligence per se, that is likely to determine how well a child learns in a practical educational setting.

Paradoxically, the auxiliary hypotheses of stage theory may have played a greater role in the development of research programs other than Piaget's own, and the applications of stage theory are frequently more sharply defined in such contexts. An example is provided by Kohlberg's (1975) theory of moral development. Kohlberg began with a stage theory formulated by Piaget himself (Piaget, 1932). He developed it essentially by putting Piaget's auxiliary hypotheses to empirical test and then replacing them by hypotheses that seemed more adequate from an empirical and a descriptive point of view. Although Kohlberg claims that there are explicit connections between the stages of his theory and Piaget's stages of concrete and formal operations, the stages themselves appear very different from Piaget's. Our concern, however, is not with a description of Kohlberg's stages or the precise nature of the difference between his approach and Piaget's. The point we wish to make is simply that Kohlberg's notion of a stage is much closer to the notion of a tightly organized, well-defined pattern of behavior and thinking than is Piaget's. The result is a theory that is much more straightforward in its application to education. In fact, it follows quite precisely from the theory what the more important conditions for the transitions between stages are—a situation quite uncharacteristic of Piaget's own theory. The major teaching technique is moral discussion between students and teachers (there are many correspondences between it and Piagetian free problem-solving activity both in its allowance of spontaneity and its element of guidance by

complete; and these in turn become necessary when the sensorimotor functions are complete. But nothing is given a priori or in an innate fashion. . . . These (structures) are all constructed and the construction—this I find to be the great mystery of the stages—becomes more and more necessary with time. The necessity is at the end of development, not at the point of departure. This then, is the model upon which we are trying to base our work and our experiments" (statement by Piaget, cited in Green, Ford, & Flamer, 1971, p. 6).

questioning). In this context, the important requirements for stage transition are exposure to (a) the next higher stage of reasoning and (b) situations posing problems and contradictions for the child's current moral structure, leading to dissatisfaction and hence the reconstruction necessary to effect a stage transition.

4. SOME SPECIFIC APPLICATIONS

This section is something of an epilogue. I have attempted to derive from Piaget's own writings a conceptually clear statement of the theory and the educational techniques that can be derived from it. Ideally, an account should now be given of how these techniques have been applied in practical situations. Unfortunately, we promptly encounter the problem of idiosyncratic interpretations of Piagetian theory that was alluded to earlier. Moreover, the nature of this interpretation interacts with the way the techniques are applied. As a result, any semblance of conceptual clarity is lost.

Therefore, this section will merely consist of a few assorted examples of the techniques being applied, the non-Piagetian components of such applications, and the difficulties encountered in implementing them. They will be drawn from the two areas where Piaget has been most influential: mathematics teaching and early childhood education (including the initial primary grades). I will finish by discussing one application, with which I have had some direct experience, in somewhat greater detail: Papert's computer-oriented LOGO project.

The reader interested in a wider variety of applications should note that there exist several collections of papers (e.g., Athey & Rubadeau, 1970; Schwebel & Raph, 1973; Varma & Williams, 1976). Most of the general introductions to Piaget contain chapters describing educational applications, the most extensive review being in the second edition of Phillips (1975). Although not a review of the literature, the recent book by Elkind (1976) is a down-to-earth discussion of specific ways in which Piagetian theory can be applied.

4.1 *Mathematics*

The hard core of Piagetian theory is replete with mathematical analogies. The structure of mathematics is one of the main sources of the auxiliary hypotheses of stage theory. His account of intellectual development is heavily influenced by his conception of the way mathematicians do their research. Therefore, it is not surprising that there are many parallels between mathematics education and Piaget's own ideas. However, the existence of such a parallel does not necessarily denote an application of Piaget's ideas to mathematics or mathematics education. Most frequently,

it simply reflects some influence that mathematics has had on Piaget's thinking.

Sometimes, a Piagetian technique and a parallel approach independently developed within mathematics come together. The most interesting example is *discovery learning*.³⁷ On the face of it, this would seem an unusually straightforward example of what we have called a free problem-solving activity. However, its development owes at least as much, and probably more, to the considerably more specific work of the mathematician George Polya (1948). There is no evidence that this was at all influenced by Piaget's work.

Discovery learning is interesting for another reason. It has, over the years, been evaluated in an unusually thorough fashion (indeed, it would seem to be the only Piagetian application for which any convincing results have been established). Much of this is reviewed in Shulman and Keislar (1966). Unlike most evaluations, much of the data comes from laboratory experiments that simulate a discovery learning situation. This enables rather precise estimates to be made of the amount of transfer of training that takes place. Such results are also easier to interpret because the task situation is defined considerably more precisely than in a naturalistic classroom setting. Some of the results are negative. Frequently, adequate transfer of training does not occur. When it does occur, the discovery learning takes an extremely long time. However, this does not prove that the Piagetian approach is wrong. Discovery learning is a compilation of techniques. The issue is which techniques and what kinds of tasks produce adequate transfer. When this is taken into account³⁸ it turns out that the necessary condition for success is that the student must actively discover the relevant principle for himself.

There are also interesting parallels between the notions underlying Piagetian constructive analysis and the notions of mathematical competence underlying the "new math" curricula. While these are merely coincidental, there has been more and more of a tendency in recent years to bring the two together. The pioneer in this regard was Lovell (1962). In recent years it has come to dominate the kind of research in mathematics education that is done in curriculum-oriented departments.³⁹ However, it is difficult to generalize from this kind of research. It is applied research, but it is concerned with highly specific problems likely to be encountered by mathematics teachers. No effort has yet been made to integrate these results and discover if they can be used to improve the curriculum.

³⁷ The best source of information on this subject is the collection of papers edited by Shulman and Keislar (1966).

³⁸ See, for example, the review of relevant research in the text by Anderson and Faust (1974).

³⁹ For a survey of Piagetian research from this point of view, see Roszkopf, Steffe, and Tarback (1971) and Steffe (1975).

Piagetian theory can also be used as a basis for training mathematics teachers. Interestingly enough, this is in accord with some of Piaget's most urgent statements regarding educational policy (1971b, 1974). One of the principal vehicles for doing this is the textbook designed for courses in teaching methods in mathematics education. Copeland (1976) has written a book of this type aimed at the elementary-school teacher. There is little about free problem-solving activities in this book (two pages are devoted to discovery learning). The emphasis is on stage theory and the simpler aspects of constructive analysis. All this is embedded in an approach to arithmetic and elementary geometry that emphasizes perceptual and manipulative activities with physical objects. However, the view the author gives of Piagetian theory is somewhat one-sided. There are at least three problems. The first is the lack of any distinction between figurative and operative knowledge. The second is a tendency to mistake auxiliary hypotheses for the hard core of the theory and to overgeneralize the range of tasks to which those hypotheses (especially the ones to do with stages) apply. The third is to confine the Piagetian view of structure with what might be called the mathematical educator's theory of structure. The standard paradigm for dealing with structural issues in the mathematics curriculum is to view mathematics as consisting of concepts that form some kind of relational structure. The important thing to note is that this is a static structure. A relation is not being viewed as something that can be done but as something that exists. In other words, what is lacking is the procedural, active element that plays such a large role in our exposition of the hard core of Piagetian theory.

All this is not meant to imply that the book lacks merit. In fact, it contains an excellent and lucid account of the relationship between the standard conception of the elementary mathematics curriculum and the more static and deterministic aspects of Piagetian theory. It is merely intended as a reasonably clear example of the kinds of changes that Piagetian theory goes through when an attempt is made to mold it into a useful and practical educational tool. However, it seems legitimate to wonder whether this book represents a successful application of Piaget's ideas. We have mentioned the necessity for a balance between freedom and constraint, in any genuine application. Might this book tilt the balance in the direction of excessive constraint?

4.2 *Early Education*

A large and highly visible part of the recent literature on Piagetian applications has dealt with the education of preschool children and those in the early primary grades. Partly because much of this work has been undertaken in the context of compensatory education programs funded by the United States Government, it has tended to take the form of specific curriculum projects. Three such projects will be considered

here—those of Lavatelli (1970, 1973), Furth (Furth, 1970; Furth & Wachs, 1974), and Kamii (Kamii & DeVries, 1977).⁴⁰ The first two are essentially attempts to form self-contained, easily disseminated products. Kamii's approach is considerably more subtle.

On the whole, there are two ways in which Piagetian theory can be applied to a curriculum project—directly and indirectly. The direct approach is the most dangerous. It should be clear by now that Piaget does not have a theory of teaching. There may be loosely defined techniques, but these are best viewed as heuristics or paradigms embedded in a theory of learning (in the broad sense of the term). Their effective implementation also requires a theory of instruction. If they are viewed as techniques that can be applied out of context, the theory of learning may be quite incompatible with the theory of instruction. In an effort to regain compatibility, the theory of learning may itself become distorted. Worst of all, these new theories may be hidden in the unstated presuppositions of the curriculum developer.

Lavatelli's curriculum is of this type. Essentially, it is a direct application of the auxiliary hypotheses of stage theory.⁴¹ A principal goal is to enable the child to reach the stage of concrete operations. The approach is to provide training in the tasks Piaget used to define the various stages and substages, with the sequence of content defined by the developmental sequence. The major problem with this kind of approach has already been discussed at length in our section on stage theory (Section 3.3). The tasks are merely indicator tasks, and teaching them does not guarantee that appropriate structures will be constructed.⁴² A number of other problems are pointed out by DeVries (in press). Essentially, there is a hidden theory of instruction. DeVries points to a number of aspects of Lavatelli's teaching methods (such as the use of reinforcement techniques and verbal prompts) that derive from the empirical tradition of North American experimental psychology and are quite inconsistent with the Piagetian notion of a free problem-solving activity. Most curious, however, is the notion that the Piagetian sequence of stages and substages defines

⁴⁰ Clearly, these projects have been collaborative enterprises. They are identified by a single individual's name partly for convenience and partly because that person has tended to be viewed as the originator of the project.

⁴¹ It is by no means the only such direct application. Another example is Weikart's Cognitively Oriented Curriculum (Weikart, Rogers, Adcock, & McClellan, 1971). However, Weikart uses Piagetian tasks in the context of a theory that is explicitly Weikart's own. An adequate discussion would take us far away from the main topic of this paper. For a useful critique of the approaches of Lavatelli and Weikart, in terms of their adequacy as applications of Piagetian theory, see DeVries (in press).

⁴² Sinclair (1971) provides the following metaphor: "Piaget's tasks are like the core sample a geologist takes from a fertile area and from which he can infer the general structure of a fertile soil; but it is absurd to hope that transplanting these samples to a field of nonfertile soil will make the whole area fertile" (p. 1).

an optimal sequence of instruction. A constructive analysis using the hard core of Piagetian theory would almost certainly yield the conclusion that this conjecture is false. For example, Piaget frequently finds stages where the child is using figurative rather than operative structures. In Lavatelli's approach, this is translated into the direct teaching of figurative knowledge. The child might thus acquire some figurative structures but these might well be useless in later stages.

Of course, these criticisms do not necessarily imply that the curriculum is ineffective. They merely imply that the issue of its effectiveness has little to do with Piagetian theory, except for the somewhat paradoxical fact that a proof of its effectiveness might cast some serious doubts on the usefulness of much of Piagetian theory! However, Lavatelli prefaces her curriculum guides with long expositions of Piagetian theory, and this forms the main explicit justification of her approach. What these criticisms also imply is that this must, in the the long run, totally confuse anybody who attempts to use or evaluate the curriculum.

The situation with Furth's curriculum is even more confusing. Given that he is the author of a widely respected and influential book (Furth, 1967) that pioneered in making many of the structural and epistemological aspects of Piaget's theory accessible to North American readers, one would expect that a book subtitled "Piaget's Theory in Practice" (Furth & Wachs, 1974) would be a reliable guide to the application of the hard core of the theory. It is therefore surprising to find reviewers with equally impeccable Piagetian credentials⁴³ describing it as presenting "a curriculum whose activities and rationale seem inconsistent with the ideas Piaget discusses in his books on education" (DeVries & Kamii, 1975).

One might cynically view this as something akin to a dispute between two rival theologians. However, Furth's curriculum is problematical in at least two respects. The first, which is one of the issues disturbing DeVries and Kamii, is that the "thinking games" that comprise the curriculum leave little opportunity for what I have called, in this paper, free problem-solving activities. They are exercises rather than games. To the extent that they are not simply physical training, they tend to consist of sequences of problems, each of which has a unique solution attainable by means of a simple algorithm. This limits the range of nontrivial hypotheses so drastically that it is difficult to see how the kind of active invention that Piaget describes in *To Understand Is to Invent* (1974) can take place.

The second problematical aspect, which might not disturb DeVries and Kamii but would certainly disturb the traditional educator and the average parent, is the fact that training in the basic skills (reading, arithmetic)

⁴³ Kamii teaches a course in early childhood education at the University of Geneva every summer.

is deliberately deemphasized. The curriculum is aimed at the early primary grades, yet training in the basic skills occupies only 15% of class time in the first grade. The notion that certain traditional tasks should be postponed is endemic to applications of Piagetian theory.⁴⁴ However, it is important to distinguish between innocuous postponements and radical postponements. I will call a postponement radical if it can be argued, either on the basis of a competing theory or on sociological grounds, that the resulting delay will harm the child. It seems reasonable to suggest that a radical postponement should only be implemented when the opposing arguments have been decisively refuted. It also seems reasonable to be concerned about what the postponed task is being replaced with. Unfortunately, all this creates a serious dilemma where Furth's curriculum is concerned. His notions of postponement are clearly radical. A non-Piagetian would argue that he has not refuted any of the arguments opposing the postponement of reading. On the other hand, at least some Piagetians would assert that he has not replaced reading with anything that Piaget would consider worthwhile.⁴⁵

Much of this confusion can be resolved by regarding the curriculum as an application of Furth's theory rather than Piaget's. To do this, it is necessary to go back to his earlier book on Piaget's theory of knowledge (Furth, 1969). This book has two components: (a) a collection of extracts from the writings of Piaget and his collaborators and (b) an extended commentary by Furth himself. There is a remarkable consistency between the notions in this commentary and the curriculum that developed several years later. However, many of these notions are Furth's rather than Piaget's. Indeed, Furth is quite explicit on this point. In his preface (p. ix) he lists the main notions in which his personal interpretations go beyond Piaget's "explicit writings." Two of these are particularly important. The first is that "the figurative aspect in language . . . is given emphasis." This—when combined with Piaget's notions regarding the necessity for operational structures to coordinate the figurative aspects of thinking and Furth's own somewhat puzzling notion that these must be formal operations—leads directly to the radical postponement of reading. The second is that "accommodation is defined as the application of a knowing

⁴⁴ "The main pre-occupation of parents on all levels of education and even of preschool familial education is that their offspring not turn out to be 'backward'. . . . A small child in nursery school must know how to read and count up to twenty at x years, while everything warns against artificially rushing. . . . And so it goes at each new stage" (Piaget, 1974, p. 82).

⁴⁵ The issue of how far Piaget himself would go in postponing reading is an open question. There are reasonable grounds for asserting that the theory implies that a child will not be able to comprehend until sufficient concrete operational structures are available to coordinate the figurative letter-sound correspondences. However, what this implies for reading depends on what theory of reading one adopts (Piaget does not provide one). The resulting postponements may be quite innocuous. In contrast, Furth believes that formal operational structures are necessary for comprehension. This accounts for the considerably more radical nature of his notions.

structure to a particular instance regardless of whether it issues in a new structure." It is difficult to say what Furth means by this. However, it is clearly different from the notion as defined in the present paper. Also, it is associated with a very different⁴⁶ definition of structure as "the general form, the interrelatedness of parts within an organized totality" (p. 265). The important point is that these notions are then substituted for Piaget's throughout the book. A precise analysis of what this does to Piaget's theory is beyond the scope of this paper. It is sufficient to point out that it seems to open up the notion, absent from Piagetian theory, that structures can somehow be enhanced by acquiring better form or better organization, and that this enhancement can occur through exercising the structures in appropriate new situations. From this perspective, Furth's thinking games begin to acquire an adequate rationale. But it is a rationale that comes from a theory that is no longer Piaget's.

The reader may have noticed by now that I have made extensive use in this section of the criticisms of Kamii and her collaborator DeVries. This is because they provide a convenient touchstone. Kamii is unique in having made a deliberate effort (for better or worse) to adopt an approach that is completely consistent with Piagetian theory. It is now time to consider what kind of touchstone this is.

Kamii's work can be divided into two phases. She began her studies while she was associated with Weikart's group at Ypsilanti, Michigan.⁴⁷ She would have seemed to have developed an independent approach while still at Ypsilanti. However, her writings of this period reflect a concern with developing a conventional disseminable curriculum (e.g., Sonquist, Kamii, & Derman, 1970). Her work since leaving Ypsilanti (Kamii & DeVries, 1977) has had a completely different emphasis. The above-mentioned criticisms of other curricula reflect this shift in approach.

This recent work has three important aspects. The first is the notion that Piaget's theory offers a tool rather than a method or set of principles that can be directly applied. However, it is not a tool that exists in ready-made form and it is not one that can be constructed by the developer of a prescriptive curriculum. It is a tool that must be constructed by classroom teachers themselves, "who experiment (perhaps in collaboration with psychologists) to find out what kinds of classroom procedures best engage children's active interests" and "who actively engage in the task of figuring out how children are thinking in particular situations and who then modify their interventions and planning on the basis of these observations" (DeVries, *in press*). The second is the notion that any

⁴⁶ It may be reasonable to argue that my exposition is overly influenced by notions deriving from artificial intelligence and the psychology of human information processing. However, it is equally reasonable to argue that Furth's exposition is overly influenced by notions deriving from Gestalt psychology.

⁴⁷ For this reason, there is a tendency to confuse the two approaches.

curriculum that prescribes fixed tasks (as both Furth and Lavatelli do) is doomed to be inconsistent with Piagetian theory. What must be done instead is to create an environment in which the teacher or the child can generate tasks as the need and opportunity arise. The third is that the main goal, as far as the child is concerned, is to enhance the formation of the structures that will ultimately characterize formal operational thinking.⁴⁸ Moreover, this is conceived in the broadest sense of the term. Rather than being confined to purely intellectual functioning, it is viewed as something that characterizes the whole child, his personality and attitudes, and the way he functions in his environment.

What comes out of all this is not a conventional curriculum. There are no workbooks, no packaged materials, no schedules.⁴⁹ It might best be viewed as a project for training preschool teachers in a fashion consistent with Piagetian theory. A brief description of how the training is done is given in Kamii and DeVries (1977). In the terminology of the present paper, the main emphasis is on the effective utilization of free problem-solving activities. Some informal constructive analysis is also brought in. However, the auxiliary hypotheses of stage theory are relegated to a minor diagnostic role. The constructive analysis would seem to be guided more by the hard-core hypotheses and heuristics.

Viewed objectively, Kamii's approach is not necessarily any better than the more direct or prescriptive uses of Piagetian theory that we have considered up to now. What is important, for the purposes of this paper, is that it demonstrates that it is possible to use Piaget's ideas, in practical educational settings in a nontrivial way, and remain completely consistent with the theory. It is unclear whether Kamii's approach is the only way to accomplish this. However, it does suggest a paradigm for achieving consistency. Also, it suggests that it is a paradigm that cannot be followed without paying a price. It involves abandoning the notion of the structured curriculum as a direct instrument of intervention in the teaching process.

4.3 *The LOGO Environment*

Papert's computer-based LOGO environment (Papert, 1972a, 1972b, 1972c; Papert & Solomon, 1972) has a number of components, of which

⁴⁸ These are preschool children that are being referred to. "For Piagetians, what is important is not how many right answers the child has learned in the past, but the extent to which past constructions (even if 'wrong') enable the child to construct knowledge in the future at ages ten, fifteen and beyond" (Kamii & DeVries, 1977).

⁴⁹ "Piaget's theory does not imply one specific 'curriculum' that can be neatly packaged and given as a cookbook to teachers. Instead, we can only provide general principles of teaching and examples of activities. . . . The Piagetian teacher is not a technician who puts children through prescribed procedures, but an autonomous professional who makes professional judgments" (Kamii & DeVries, 1977).

the computer itself is the least important.⁵⁰ There is a programming language (LOGO) that is simple enough for a child to use in writing programs. There is a computer terminal that a child can use to input a program into the computer and observe what it does when it is executed. There are also several computer-controlled devices, which I will collectively (though slightly inaccurately) refer to as "turtles," whose behavior is determined by the instructions the child gives to the computer in a LOGO program. For example, there is a floor turtle that can roam around the room, there is a graphics turtle consisting of a cathode-ray screen on which animated pictures can be "drawn," and there is a music box that plays music composed by the child. Finally, there are a number of things that are *not* controlled by the computer. In addition to the kinds of things normally found in a classroom (pencils, paper, paints, etc.), this includes equipment with which a child can practice activities, such as juggling and balancing, that demand the coordination of physical skills.

On the surface, all this may seem to have little to do with Piaget. Indeed, no description of LOGO published to date makes any direct claim that Piagetian theory is being applied. Why, then, is there any point in discussing it in this paper? There are two reasons. The first is that there are circumstantial grounds for searching for an influence of some kind, since Papert spent several years working at Piaget's laboratory and has maintained reasonably close contact since leaving.⁵¹ The second is that the literature on LOGO that currently exists does not provide a specific rationale or a clear set of goals. While it is a perfectly well defined environment, it is currently unclear what should be done with it. It makes sense, to some extent, as a way of teaching some useful aspects of computer science.⁵² However, it is clear even from the published literature (and considerably clearer if one talks to the few individuals who know how to use the LOGO environment effectively with children) that there is considerably more to it than that. It seems reasonable to inquire whether Piagetian theory provides any clues as to what this might be.

The relationship between LOGO and Piagetian theory can be approached at two levels. At the more superficial level, it is clear from examining the published literature on LOGO that there is a fairly close

⁵⁰ It is unimportant in two senses. The first is that it does not matter what kind of computer is used, as long as it is capable of creating and maintaining the environment. The second is that the child does not need to know anything about the computer, except that it is a thing that faithfully obeys instructions (even wrong ones).

⁵¹ For some articles by Papert dealing with issues of Piagetian theory (unrelated to LOGO) see Volumes 11, 15, and 22 of *Études d'Épistémologie Génétique* (Gréco, Grize, Papert, & Piaget, 1960; Apostel, Grize, Papert, & Piaget, 1963; Cellérier, Papert, & Voyat, 1968). For a very brief comment on LOGO by Piaget himself, see Piaget (1974, p. 7).

⁵² Indeed, there exists what might be called a "computer science approach" to using LOGO (e.g., Feurzeig & Lukas, 1972) that is somewhat different from Papert's.

correspondence between the Piagetian free problem-solving "technique" and the way Papert feels that the LOGO environment should be used. To quote from the introduction to *Teaching Children Thinking* (Papert, 1972b):

The phrase "technology and education" usually means inventing new gadgets to teach the same old stuff in a thinly disguised version of the same old way. . . . The purpose of this essay is to present a grander vision of an educational system which technology is used not in the form of machines for processing children, but as something the child will learn to manipulate, to extend, to apply to projects, thereby gaining a greater and more articulate mastery of the world, a sense of the power of applied knowledge and a self-confidently realistic image of himself as an intellectual agent. . . . I claim that computation is by far the richest known source of these ingredients. We can give children unprecedented power to invent and carry out exciting projects by providing them with access to computers, with a suitably clear and intelligible programming language and with peripheral devices capable of producing on-line real-time action. (p. 245)

Later on in the same paper, there is an even more explicit reference to the ideas we have discussed in Section 3: "The child does not wait with a virginally empty mind until we are ready to stuff it with a statistically validated curriculum. He is constantly engaged in inventing theories about everything, including himself, schools and teachers" (p. 246). There is even the notion, which we have already encountered in our discussion of Kamii's work, that rigidly defined tasks are detrimental to this kind of learning. Referring to his experiences in teaching seventh-grade children, Papert makes the point that "the most exciting single aspect of the experiment was that most of the children acquired the ability and motivation to work on *projects* that extend in time over several days, or even weeks. This is in marked contrast with the style of work in most mathematics classes where the children work on problems of much shorter duration and more rigidly defined by the teachers. Our goal is to achieve a work style similar to that of the better art classes" (p. 252).

All this, and similar statements elsewhere, can be taken to imply that the best way to use the LOGO environment (from Papert's point of view) is to give the child the opportunity to indulge in free problem-solving activities. The goal is to improve "the child's ability to articulate the working of his own mind and particularly the interaction between himself and reality in the course of learning and thinking." The approach is to follow the model (considered in Sec. 3) of the child as a researcher and the teacher as research director. The child selects a project and is free to do anything he wishes to accomplish it, subject to quite explicit constraints imposed by the LOGO environment. It need not necessarily involve the computer. It may involve developing a physical skill, such as juggling. The important thing is that it is a project in which the child has to worry about the procedures he or she is using. A plan must be conceived, executed,

and *debugged*.⁵³ Debugging is particularly important. The environment ensures that a child will obtain automatic and quite subtle feedback regarding the errors in his or her current plan (see the discussion at the end of Sec. 3.2). Successful debugging can be accomplished only by deliberately reflecting upon what the plan really consists of and what there is about it that is causing the errors.

From what has been considered so far, it seems reasonable to conclude that Papert has in mind something akin to the instructional approach advocated by Piaget. However, there is a deeper issue involved. It is no use teaching trivial things in this fashion. What are all these interesting projects actually accomplishing? To answer this question, it is necessary to go beyond the published literature on the LOGO environment and consider what kinds of structures might be formed through the activity of inventing computer programs that do interesting things.

It might be asserted that what is being learned is the set of basic features of a programming language. However, this is insufficient to explain the LOGO environment. Also, such knowledge is trivial from a Piagetian point of view. It is simply figurative knowledge coordinated by low-level operative structures. There may be arguments against this position, but it must be borne in mind that I am attempting a Piagetian analysis. Moreover, it is not as controversial as it may seem. Something that is trivial is not necessarily unimportant. For example, reading and writing are essential skills for functioning in our society but their basic elements (from a Piagetian point of view) are low-level structures of this same figurative kind. Also, it is not inconsistent with the point of view of practical computer programmers. What is important is not the syntax and semantics of the programming language *per se*, but the interesting and important things that can be accomplished through its use.

For a satisfactory answer to our question, it is necessary to turn to the hard core of Piaget's theory. The important thing is that computer programs are structures, in the strict sense of the definition in Section 2. Moreover, they are structures that function and they have an inherent hierarchical organization, in the sense that there is a main program that coordinates the functioning of subprograms and these in their turn can be viewed as programs with their own subprograms, and so on. In other words, computer programs are structures that coordinate other structures. While a computer program does not usually grow (especially one that a normal child is likely to write), an analogous notion can be obtained by considering the inevitable debugging process. The result of such a

⁵³ A technical term in computer programming for the iterative process of discovering why a program does not work. Anybody who has ever written a nontrivial computer program should be familiar with the peculiarly self-motivating nature of this process.

process is a sequence of programs. Each program is closely related to the preceding program in the sequence, but each "works" better than the one that preceded it. In Piagetian terms, one might say that these programs are acquiring (through accommodation) a progressive ability to assimilate reality.

What is important about all this is that it follows from Piagetian theory that the child who has invented the computer program must have, in his or her head, some structure that models the *functioning* of the computer program. As he or she goes through the debugging process, this structure must change in the same way as the program changes. Thus, the child's structures will accommodate and assimilate reality as a result of the debugging process. Moreover, he or she is much more likely than usual to be aware of the assimilation and accommodation that are actually going on.

Up to now, we have only considered what happens when a child writes a single program. It is also necessary to consider what happens when he or she invents and debugs many programs over a long period of time. It seems reasonable to assert that the child will develop general heuristics for organizing programs so that it is easy to accomplish desirable goals, and so that the programs themselves are easier to debug. These general heuristics (Papert calls them "powerful ideas") are at the heart of LOGO. The basic hypothesis is that they are not simply ways of organizing computer programs. They can be utilized to organize all kinds of procedural knowledge (which is why nonprogramming activities with a procedural orientation, such as learning juggling, are present in the LOGO environment). They are also at the heart of Piagetian theory. They must reflect the functioning of very high level logico-mathematical structures. According to Piaget, it is the existence of such structures that is responsible, in the long run, for the higher levels of intelligent human activity.

It would be possible to write considerably more about the LOGO environment in this vein, especially about what might happen over a long period of time. However, enough has been said to give a sketch of how Piagetian theory can be used to "explain" LOGO. Does this help? The answer depends on how much Piagetian theory the reader is willing to accept. However, it seems reasonable to say that such an exercise does serve two useful functions. The first is that it shifts the focus of LOGO away from the computer. In Piagetian terms, everything is there because it enhances the development of logico-mathematical structures. The turtles are not merely interesting toys. They are there to give the structures something concrete to coordinate. The programming language is there because programs can be written in it, and programs give objective embodiment to structural notions. The computer is there simply to provide a self-correcting environment in which structures can function and change. The second is that it provides a novel way of looking at Piagetian theory and the issue of how it should be applied. Most important, it suggests that

the kind of procedural epistemology that arises from computer science may provide some insight into the problem of what structures really consist of and how they change.

5. SOME CONCLUDING COMMENTS

Three main issues have been dealt with in this paper: the nature of the theory itself, the misconceptions that have arisen, and the educational implications that can be legitimately drawn. It is now necessary to consider whether we have made any progress in clarifying them.

As far as the theory itself is concerned, our main contribution has been to provide a number of advantages over more conventional ways of describing Piaget's work. It enables much to be said in a concise fashion. It puts the more ephemeral aspects of the theory in their proper place. It enables at least some hints to be given of the subtlety of the structural and epistemological notions and of the basic role they play in the theory. Most important, for the purposes of this discussion, it provides a convenient conceptual vocabulary for discussing the other two issues.

The various misconceptions that we have considered usually arise either from the mistaking the auxiliary hypotheses for the hard core or from making auxiliary hypotheses that are inconsistent with the hard core. Of course, there is nothing wrong in disagreeing with Piaget. The problems arise when the disagreement is not explicitly stated and its consequences are not analyzed in sufficient depth. The result is a new research program with a hard core that is likely to be impoverished and may be based on a set of mutually contradictory hypotheses. The pervasiveness of this practice is one explanation for the existence of such a large number of weak theories, all different yet all attributed to Piaget.

Such misconceived "Piagetian" research programs are very likely to occur in education, and our analysis suggests a logical reason for this. The notion that something will be effective for instructional purposes can be viewed as an auxiliary hypothesis in some kind of research program. As we have noted, however, Piaget's own research program does not contain a theory of instruction. Hence, it is *always* necessary to add to the hypotheses that already exist. Owing to the peculiarly interconnected nature of Piaget's theory, there is a high probability that the investigator who is primarily interested in developing practical techniques with a short-term payoff will develop a set of auxiliary hypotheses that are inconsistent with the hard core.

Finally, let us turn to the legitimate implications of the theory and the question of how they can be put into practice. First of all, it is necessary to dispose of an apparent contradiction. We have gone to great pains to assert that Piaget does not have a theory of instruction. In Section 3, however, some techniques were described that are not only consistent with

the theory but were suggested by Piaget himself. Do these not constitute a theory of instruction? They do not, because they are not practical techniques. They are too loosely defined to be implementable without a great deal of elaboration. In terms of our vocabulary, they are heuristics rather than auxiliary hypotheses. It is their elaboration, together with the hypotheses that must be built up around them to develop something practical and effective, that constitutes a theory of instruction.

With this out of the way, it is possible to say what a legitimate educational implication consists of. It is something in the theory that has heuristic value and can be developed into a theory of instruction whose hard core includes the hard core of the Piagetian research program. This new theory need not necessarily make use of the techniques that Piaget himself suggests. What is important is the consistency of the auxiliary hypotheses with the hard core of Piaget's own theory. It should be clear, by now, that the development of such a theory is not a simple task.

In general, a psychological theory serves education as a source of good ideas. There are two ways of using such ideas—either consistently with the theory or inconsistently with it. In this paper, the emphasis has been on the first approach. This is partly because so many educators have unsuccessfully tried to follow it. The main reason, however, is that this has been a paper on Piaget's theory rather than somebody else's. The second path can only be *deliberately* followed by fitting the ideas into another psychological theory. On the other hand, it would be unfortunate if the reader were left with the impression that the second path is an undesirable one to pursue. The only requirement is that the person following it should be willing to take credit for his own ideas, and not attribute them to Piaget. There is nothing sacred about Piaget's theory. A research program with a different hard core might yield much better ideas, yet still be amenable to many of Piaget's.

It is difficult to say which path is superior. Indeed, it seems reasonable to assert that both are necessary for the development of a successful science of education. The problem with Piaget's theory is that it is too good. It is the only well-worked-out psychological theory that takes knowledge seriously. Without competitors, there is a danger that it will turn into a creed. It is not through dogma but by the clash of rival research programs that science flourishes.

REFERENCES

- Aebli, H. *Didactique psychologique*. Neuchatel: Delacheux et Niestle, 1951.
- Anderson, R. C., & Faust, G. W. *Educational psychology: The science of instruction and learning*. New York: Dodd, Mead, 1974.
- Apostel, L., Grize, J. B., Papert, S., & Piaget, J. *La filiation des structures*. Paris: Presses Universitaires de France, 1963.

- Arbib, A., & Mannes, E. *Arrows, structures and functors: The categorical imperative*. New York: Academic Press, 1975.
- Athey, I. J., & Rubadeau, D. O. (Eds.). *Educational implications of Piaget's theory*. Waltham, Mass.: Ginn-Blaisdell, 1970.
- Beilin, H. Developmental stages and developmental processes. In D. R. Green, M. Ford, & G. B. Flamer (Eds.), *Measurement and Piaget*. New York: McGraw-Hill, 1971. (a)
- Beilin, H. The training and acquisition of logical operations. In M. F. Roszkopf, L. P. Steffe, & S. Tarback (Eds.), *Piagetian cognitive-development research and mathematical education*. Reston, Va.: National Council of Teachers of Mathematics, 1971. (b)
- Berlyne, D. E. *Structure and direction in thinking*. New York: McGraw-Hill, 1965.
- Beth, E. W., & Piaget, J. *Mathematical epistemology and psychology* (W. Mays, trans.). New York: Gordon & Breach, 1966. (Originally published, 1961.)
- Bruner, J. S. *Process of education*. New York: Vintage Books, 1960.
- Bruner, J. S., Goodnow, J., & Austin, G. *A study of thinking*. New York: Wiley, 1956.
- Bruner, J. S., Olver, R. R., & Greenfield, P. M. *Studies in cognitive growth*. New York: Wiley, 1966.
- Cellérier, G., Papert, S., & Voyat, G. *Cybernétique et épistémologie*. Paris: Presses Universitaires de France, 1968.
- Chomsky, N. *Syntactic structures*. The Hague: Mouton, 1957.
- Copeland, R. W. *Mathematics and the elementary teacher* (3rd ed.). Philadelphia: Saunders, 1976.
- DeVries, R. Early education: Applications versus implications. In J. M. Gallagher & J. A. Easley (Eds.), *Knowledge and development* (Vol. 2). New York: Plenum, in press.
- DeVries, R., & Kamii, C. Review of *Thinking goes to school* by H. Furth and H. Wachs. *Young Children*, 1975, **30**, 378-379.
- Elkind, D. *Child development and education*. New York: Oxford University Press, 1976.
- Elliott, C. The measurement of development. In V. P. Varma & P. Williams (Eds.), *Piaget, psychology and education: Papers in honor of Jean Piaget*. Itasca, Ill.: Peacock, 1976.
- Engelmann, S. Does the Piagetian approach imply instruction? In D. R. Green, M. Ford, & G. B. Flamer (Eds.), *Measurement and Piaget*. New York: McGraw-Hill, 1971.
- Feigenbaum, E. A., & Feldman, J. G. *Computers and thought*. New York: McGraw-Hill, 1963.
- Feurzeig, W., & Lucas, G. LOGO—A programming language for teaching mathematics. *Educational Technology*, March 1972.
- Flavell, J. H. *The developmental psychology of Jean Piaget*. Princeton, N.J.: Van Nostrand, 1963.
- Fogelman, K. R. *Piagetian tests for the primary school*. London: National Foundation for Research in England and Wales, 1970.
- Furth, H. G. *Piaget and knowledge—Theoretical foundations*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.
- Furth, H. G. *Piaget for teachers*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- Furth, H. G., & Wachs, H. *Thinking goes to school—Piaget's theory in practice*. New York: Oxford University Press, 1974.
- Ginsburg, H. *The myth of the deprived child: Poor children's intellect and education*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

- Ginsburg, H., & Oppen, S. *Piaget's theory of intellectual development: An introduction*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.
- Gréco, P., Grize, J. B., Papert, S., & Piaget, J. *Problèmes de la construction du nombre*. Paris: Presses Universitaires de France, 1960.
- Green, D. R., Ford, M., & Flamer, G. B. (Eds.). *Measurement and Piaget*. New York: McGraw-Hill, 1971.
- Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
- Hunt, J. McV. *Intelligence and experience*. New York: Ronald Press, 1961.
- Inhelder, B. *Diagnosis of reasoning in the mentally retarded* (W. B. Stephens, trans.). New York: John Day, 1968. (Originally published, 1943.)
- Inhelder, B., & Piaget, J. *The early growth of logic in the child: Classification and seriation*. New York: Harper & Row, 1964. (Originally published, 1959.)
- Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (A. Parson & S. Milgram, trans.). New York: Basic Books, 1958. (Originally published, 1955.)
- Inhelder, B., Sinclair, H., & Bovet, M. *Learning and the development of cognition* (S. Wedgwood, trans.). Cambridge, Mass.: Harvard University Press, 1974.
- Isaacs, N. *A brief introduction to Piaget*. New York: Schocken Books, 1974.
- Kamii, C., & Derman, L. The Engelmann approach to teaching logical thinking. In D. R. Green, M. Ford, & G. B. Flamer (Eds.), *Measurement and Piaget*. New York: Graw-Hill, 1971.
- Kamii, C., & DeVries, R. Piaget for early education. In M. C. Day & R. K. Parker (Eds.), *Preschool in action* (2nd ed.). Boston: Allyn & Bacon, 1977.
- Kessen, W., & Kuhlman, C. (Eds.). Thought in the young child. *Monographs of the Society for Research in Child Development*, 1973, 27(2).
- Koch, S. (Ed.). *Psychology, a study of a science* (6 vols.). New York: McGraw-Hill, 1959-1963.
- Kohlberg, L. The cognitive-developmental approach to moral education. *Phi Delta Kappan*, 1975, 56, 670-677.
- Kuhn, T. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.
- Lakatos, I. Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge*. Cambridge, England: Cambridge University Press, 1972.
- Laudan, L. *Progress and its problems*. Berkeley: University of California Press, 1977.
- Lavatelli, C. *Piaget's theory applied to an early childhood curriculum*. Boston: American Science and Engineering, 1970.
- Lavatelli, C. *Teacher's guide to accompany early childhood curriculum—A Piaget program*. Boston: American Science and Engineering, 1973.
- Lovell, K. *The growth of basic mathematical and scientific concepts in children* (2nd ed.). London: University of London Press, 1962.
- Luce, D., Bush, R., & Galanter, E. (Eds.). *Handbook of mathematical psychology* (3 vols.). New York: Wiley, 1962-1965.
- Lunzer, E. A. *Recent studies in Britain based on the work of Jean Piaget*. London: National Foundation for Educational Research, 1960.
- MacLane, S. *Categories for the working mathematician*. New York: Springer-Verlag, 1972.
- Miller, G. A., Galanter, E., & Pribram, K. *Plans and the structure of behavior*. New York: Holt, 1960.
- Modgil, S. *Piagetian research: A handbook of recent studies*. Windsor, England: National Foundation for Educational Research, 1974.

- National Froebel Foundation. *Some aspects of Piaget's work*. London: Author, 1955.
- Newell, A., Shaw, J. C., & Simon, H. A. Elements of a theory of human problem solving. *Psychological Review*, 1958, **65**, 151-166.
- Newell, A., & Simon, H. A. *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.
- Osherson, D. N. *Logical abilities in children*. Potomac, Md.: Lawrence Erlbaum & Associates, 1974.
- Papert, S. A computer laboratory for elementary schools. *Computers and Automation*, 1972, **21**(6). (a)
- Papert, S. Teaching children thinking. *Programmed Learning and Educational Technology*, 1972, **9**(5), 245-255. (b)
- Papert, S. Teaching children to be mathematicians versus teaching about mathematics. *International Journal of Mathematics Education and Science Technology*, 1972, **3**, 249-262. (c)
- Papert, S., & Solomon, C. Twenty things to do with a computer. *Educational Technology*, April 1972, pp. 9-18.
- Phillips, J. L. *The origins of intellect: Piaget's theory* (2nd ed.). San Francisco: Freeman, 1975.
- Piaget, J. *The language and thought of the child* (M. Worden, trans.). New York: Harcourt, Brace & World, 1926. (Originally published, 1923.)
- Piaget, J. *Judgment and reasoning in the child* (M. Worden, trans.). New York: Harcourt, Brace & World, 1928. (Originally published, 1924.)
- Piaget, J. *The child's conception of the world* (J. Tomlinson & A. Tomlinson, trans.). New York: Harcourt, Brace & World, 1929. (Originally published, 1926.)
- Piaget, J. *The child's conception of physical causality* (M. Worden, trans.). New York: Harcourt, Brace & World, 1930. (Originally published, 1927.)
- Piaget, J. *The moral judgment of the child* (M. Worden, trans.). New York: Harcourt, Brace & World, 1932.
- Piaget, J. *Classes, relations et nombres*. Paris: Vrin, 1942.
- Piaget, J. *Traite de logique*. Paris: Colin, 1949.
- Piaget, J. *Introduction à l'épistémologie génétique* (3 vols.). Paris: Presses Universitaires de France, 1950.
- Piaget, J. *The psychology of intelligence* (M. Piercy & D. E. Berlyne, trans.). London: Routledge & Kegan Paul, 1950. (Originally published, 1947.)
- Piaget, J. *Play, dreams, and imitation in childhood* (C. Gattegno & F. M. Hodgson, trans.). New York: W. W. Norton, 1951. (Originally published, 1945.)
- Piaget, J. *Essai sur les transformations des opérations logiques*. Paris: Presses Universitaires de France, 1952.
- Piaget, J. *The origins of intelligence in children* (M. Cook, trans.). New York: International Universities Press, 1952. (Originally published, 1936.)
- Piaget, J. How children form mathematical concepts. *Scientific American*, 1953, **189**(5), 74-79.
- Piaget, J. *The construction of reality in the child* (M. Cook, trans.). New York: Basic Books, 1954. (Originally published, 1937.)
- Piaget, J. The child and modern physics. *Scientific American*, 1957, **196**(3), 46-51. (a)
- Piaget, J. *Logic and psychology*. New York: Basic Books, 1957. (b)
- Piaget, J. Review of *Studies in cognitive growth* by J. S. Bruner, R. Olver, and P. Greenfield. *Contemporary Psychology*, 1967, **12**, 532-533.
- Piaget, J. *The child's conception of time* (A. J. Pomerans, trans.). London: Routledge & Kegan Paul, 1969. (Originally published, 1946.)

- Piaget, J. *The child's conception of movement and speed* (G. E. T. Holloway & M. J. Mackenzie, trans.). London: Routledge & Kegan Paul, 1970. (Originally published, 1946.)
- Piaget, J. *Genetic epistemology*. New York: Norton, 1971. (a)
- Piaget, J. *Science of education and the psychology of the child*. New York: Viking, 1971. (b)
- Piaget, J. *Biology and knowledge*. Chicago: University of Chicago Press, 1971. (Originally published, 1967.)
- Piaget, J. *Structuralism* (C. Maschler, trans.). New York: Basic Books, 1971. (Originally published, 1968.)
- Piaget, J. *The principles of genetic epistemology* (W. Mays, trans.). London: Routledge & Kegan Paul, 1972. (Originally published, 1970.)
- Piaget, J. *To understand is to invent—The future of education*. New York: Viking, 1974. (A translation of two works written for UNESCO in 1948 and 1971.)
- Piaget, J. *The grasp of consciousness*. Boston: Harvard University Press, 1976. (Originally published, 1974.)
- Piaget, J., & Inhelder, B. *The child's conception of space* (F. J. Langdon & J. L. Lunzer, trans.). London: Routledge & Kegan Paul, 1956. (Originally published, 1948.)
- Piaget, J., & Inhelder, B. *Psychology of the child* (H. Weaver, trans.). London: Routledge & Kegan Paul, 1969. (Originally published, 1967.)
- Piaget, J., & Inhelder, B. *Mental imagery in the child*. London: Routledge, Kegan, Paul, 1971. (Originally published, 1966.)
- Piaget, J., & Inhelder, B. *Memory and intelligence* (A. J. Pomerans, trans.). New York: Basic Books, 1972. (Originally published, 1968.)
- Piaget, J., Inhelder, B., & Szemirska, A. *The child's conception of geometry* (E. A. Lunzer, trans.). New York: Basic Books, 1960. (Originally published, 1948.)
- Piaget, J., & Szemirska, A. *The child's conception of number* (C. Gattegno & F. M. Hodgson, trans.). New York: Humanities Press, 1952. (Originally published, 1941.)
- Pinard, A., & Larendau, M. A scale of mental development based on the theory of Piaget. *Journal of Research on Science Teaching*, 1964, **2**, 253-260.
- Polya, G. *How to solve it*. Princeton, N.J.: Princeton University Press, 1948.
- Rasch, G. An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 1966, **19**, 49-57.
- Ripple, R. E., & Rockcastle, V. N. Piaget rediscovered: Selected papers from a conference on cognitive studies and curriculum development. *Journal of Research in Science Teaching*, 1964, **2**(3).
- Roszkopf, M. F., Steffe, L. P., & Tarback, S. (Eds.). *Piagetian cognitive-development research and mathematical education*. Reston, Va.: National Council of Teachers of Mathematics, 1971.
- Schwebel, M., & Raph, J. (Eds.). *Piaget in the classroom*. New York: Basic Books, 1973.
- Shulman, L. S., & Keislar, E. R. (Eds.). *Learning by discovery: A critical appraisal*. Chicago: Rand McNally, 1966.
- Sigel, I. E., & Hooper, F. (Eds.). *Logical thinking in children*. New York: Holt, 1968.
- Sinclair, H. Piaget's theory of development: The main stages. In M. F. Roszkopf, L. P. Steffe, & S. Tarback (Eds.), *Piagetian cognitive-development research and mathematical education*. Reston, Va.: National Council of Teachers of Mathematics, 1971.
- Sonquist, H., Kamii, C., & Derman, L. A Piaget-derived preschool curriculum. In I. J. Athey & D. O. Rubadeau (Eds.), *Educational implications of Piaget's theory*.

- Waltham, Mass.: Ginn-Blaisdell, 1970.
- Steffe, L. P. (Ed.). *Research on mathematical thinking of young children*. Reston, Va.: National Council of Teachers of Mathematics, 1975.
- Suppes, P. *Sets and numbers*. New York: Random House, 1969. (Originally published, 1960.)
- Terman, L. M. *The measurement of intelligence*. Boston: Houghton-Mifflin, 1916.
- Tuddenham, R. D. Theoretical regularities and individual idiosyncracies. In D. R. Green, M. Ford, & G. B. Flamer (Eds.), *Measurement and Piaget*, New York: McGraw-Hill, 1971.
- Varma, V. P., & Williams, P. *Piaget, psychology and education: Papers in honor of Jean Piaget*. Itasca, Ill.: Peacock, 1976.
- Wallace, J. G. *Concept growth and the education of the child*. Slough, England: National Foundation for Educational Research, 1965.
- Weikart, D., Rogers, L., Adcock, C., & McClellan, D. *The cognitively oriented curriculum*. Urbana, Ill.: Educational Resources Information Center, National Association for the Education of Young Children, 1971.

Chapter 5

PSYCHOANALYSIS AND AMERICAN ELEMENTARY EDUCATION

by

PATRICK SUPPES AND HERMINE WARREN

*Institute for Mathematical Studies
in the Social Sciences
Stanford University*

BEFORE Freud, mental disturbances were generally understood to be organic in origin. Psychoanalysis grew out of discoveries made by Freud that these disturbances were often better treated by psychological means than by medical approaches. Because he developed his therapeutic method and theory simultaneously, the term *psychoanalysis* has since come to refer to a loosely knit theory of mental processes, to the techniques by which these processes are investigated, e.g., dream analysis, and to a therapeutic procedure.

In assessing the impact that psychoanalysis has had on American elementary education, we have restricted our analysis in certain ways. Thus, in the discussion to follow, it is primarily the theory of psychoanalysis that we assess for its impact on education, with other issues brought in only occasionally. We also concentrate on what is generally referred to as "orthodox" psychoanalytic thinking; psychoanalytic schools such as those represented by Alfred Adler are given only minimal consideration.

Many people made helpful suggestions or otherwise provided assistance during the preparation of this manuscript. We would like to thank M. Almy, B. Bettelheim, J. C. Burnham, H. H. Clark, L. A. Cremin, L. J. Cronbach, A. Freud, N. L. Gage, N. G. Hale, Jr., E. R. Hilgard, D. G. Kanerva, L. Messing, H. B. Meyer, J. H. Nance, I. Paret, P. Paret, D. C. Phillips, R. R. Sears, M. K. Shaw, L. M. Stoltz, D. Suppes, and R. M. Suzman.

The sheer volume of the education literature makes it necessary to limit as well that part of it we have been able to consider in any detail. For this reason we restrict our study to the elementary education of normal children and do not address ourselves to the various counseling services, child guidance clinics, and residential schools that serve children with special problems. We also limit our assessment to influence in the United States, to the neglect of the important impact of psychoanalysis on educational work done in Austria, England, Germany, Israel, and other places.¹

The period we cover begins with the turn of the century, the time when references to psychoanalysis began to appear in the American literature, and ends in 1957. The year 1957 marks the symbolic death of the progressive education movement as well as the launching of Sputnik and on both these counts is broadly thought of as the end of an era in education.

The chapter is organized along the following lines. In the first section, a brief synopsis of some key works on the history of psychoanalysis in medicine and psychology is presented for the perspective it provides of the impact of psychoanalysis on the professions in the United States. In the second section, we provide a review of the psychoanalytic literature concerning topics of relevance to ordinary education; this section begins with a précis of Anna Freud's *Psychoanalysis for Parents and Teachers*, a work now considered classic for its clear presentation of psychoanalytic ideas.

The third section consists of a presentation of the evidence of impact of psychoanalytic thinking on education, restricted primarily to textbooks and journals, although some institutional surveys are reviewed; in general we have not evaluated the adequacy of presentation of psychoanalytic ideas but, rather, concentrate on the assessment of magnitude of impact. In the fourth section, we make a brief conceptual comparison of Freud, Dewey, and Thorndike, with an emphasis on their differing conceptions of human nature. The fifth section summarizes our conclusions.

1. BRIEF HISTORICAL OVERVIEW OF PSYCHOANALYSIS IN AMERICAN MEDICINE AND PSYCHOLOGY

Despite its European origins, psychoanalysis has had a greater impact on American culture than on any other. To provide at least a partial framework for comparing developments in education, we discuss in this section the general way in which psychoanalysis affected the professions of medicine and psychology in America.

¹ Two histories of the relation between psychoanalysis and education that emphasize the impact of Freud on education in Europe are Hoffer (1945) and Ekstein and Motto (1969); a discussion of kibbutz education from a psychoanalytic point of view is to be found in Bettelheim (1969).

1.1 Medicine

After an early abstract by William James in 1894 of the Breuer-Freud paper "On the Psychical Mechanism of Hysterical Phenomena: Preliminary Communication," Freud's beginning efforts received scant attention in the American medical press. Not until about 1905 did American physicians begin to take serious note of his work.

The influence of psychoanalysis on the medical profession has been charted in a monograph by Burnham (1967), covering the period 1894 to 1918; the history of psychoanalysis in medicine through the post World War II years is taken up by Oberndorf (1953), a past president of the American Psychoanalytic Association. These, as well as other works, show how the transmission of psychoanalytic ideas in American medicine depended largely on the personal efforts of a very small number of physicians and not on the scientific literature on psychoanalysis.

During the first decade of the century, nearly all the American physicians interested in psychoanalysis were located on the East Coast, specifically in the cities of Boston, New York, and Baltimore-Washington. Early interest of this kind may be dated to about 1905, when the Boston neurologist Morton Prince organized some weekly discussions at his home to consider issues of psychopathology. As a result of these meetings, in 1906 he founded the *Journal of Abnormal Psychology*, which from the beginning gave coverage to psychoanalysis. Members of this early Boston group included, at least from time to time, Ernest Jones, a neurologist later to become Freud's biographer; the psychologist G. Stanley Hall; and James J. Putnam, professor of neurology at Harvard. Burnham suggests that the professional standing of such men as Prince and Putnam was an important factor in gaining a hearing for psychoanalysis.

Adolf Meyer in New York was also important in spreading psychoanalytic ideas. Meyer was a Swiss immigrant and psychiatrist associated with the New York state hospital system whose early interest in psychoanalysis was reflected in the 1905 and 1906 issues of the *Psychological Bulletin* in which he wrote rather lengthy reviews of Jung's word-association experiments. Because of Meyer, along with his colleague August Hoch, young physicians in the New York state hospital system became aware of psychoanalysis.

Also in New York was Frederick Peterson, a distinguished psychiatrist and coauthor with Jung on aspects of the association experiments. Another important figure was A. A. Brill, who later translated Freud's works into English and was responsible for converting Smith Ely Jelleffe, a distinguished neurologist and editor of the *Journal of Nervous and Mental Diseases*, to psychoanalysis. This conversion was an important one, since by 1913, Jelleffe and his close friend William A. White had started the

Psychoanalytic Review, the first psychoanalytic journal in English. White was then the superintendent of the Government Hospital for the Insane in Washington, and it was through his influence that the younger psychiatrists on his staff, like those on Meyer's staff in New York, were introduced to psychoanalysis. We discuss later how White also played a significant role in bringing psychoanalytic ideas into education through the mental hygiene movement.

By 1911, the New York Psychoanalytic Society and the American Psychoanalytic Association, the first psychoanalytic professional associations, had been formed. By World War I, psychoanalysis was recognized as a subspecialty in psychiatry, though not without opposition. Oberndorf, whose largely autobiographical account includes this early period as well as later years, provides insight into the storms created by the dissensions of such psychoanalysts as Alfred Adler, Karen Horney, and Harry Stack Sullivan.

Generally speaking, these works emphasize how the American temper, as reflected in both medical and cultural attitudes, encouraged a highly pragmatic outlook, and the value of psychoanalysis was to a large extent judged by its value in treatment. The sexual emphasis in psychoanalysis seems to have been the cause of most of the explicit opposition.

1.2 Psychology

To some extent the early history of psychoanalysis in psychology is related to its history in medicine in that many of the same people, e.g., William James, G. Stanley Hall, and James J. Putnam, were responsible for bringing psychoanalysis to the early attention of both professions.

Shakow and Rapaport (1964) have chronicled in considerable detail the impact of psychoanalysis on American psychology. They discuss at length how psychology's commitment to a program of scientific experimentation, with detailed control of variables, made the field largely unreceptive to psychoanalytic thinking, which relied for its development on the empirical but nonexperimental case-study method.

These preoccupations are reflected in the responses of important psychologists to Freud; in a four-part article entitled "Contributions of Freudism to Psychology," published in the 1924 issue of the *Psychological Review*, for example, statements are to be found by L. L. Thurstone, James H. Leuba, K. S. Lashley, and Joseph Jastrow. All their remarks are worth quoting, but the statements by Thurstone and Lashley are especially representative of the thinking of psychologists.

Thurstone is sympathetic to psychoanalysis and emphasizes that the theory deals with questions that are of considerably broader interest than the questions ordinarily dealt with in academic psychology. He makes this more explicit by emphasizing that academic psychology has primarily dealt with momentary mental states, "while the new psychology deals

mostly with the *expression of basic and permanent human wants*" (p. 177). At the same time, he criticizes psychoanalysis for the absence of scientific methods of the sort used in academic psychology. The optimistic advice he gives in his final paragraph has unfortunately not been followed with any great success.

Let us turn to the effects which the psychoanalysts and the psychiatrists have used on an empirical basis in their medical practice and apply to these phenomena the methods of controlled scientific experimentation. It is certain that both medical practice and psychology as a science will profit from such a venture. (p. 183)

In contrast, Lashley is totally negative in his evaluation of psychoanalysis—though in a way that appears rather naive today. He seems to feel that, by looking to physiological or neurological theory and findings, it should be possible to eliminate the need for the kind of psychological constructs introduced in psychoanalysis; in fact, Lashley entitles his contribution "Physiological Analysis of the Libido." The kind of program that Lashley outlines is no more possible of execution now than it was at the time of its writing, for psychoanalytic theory or for any other substantial domain of psychological theory.

It is ironic that to an extent probably much truer than in the case of general medicine, Freud's ideas were well received by a large number of prominent American psychologists in the first half of the century, even though they did not use the ideas in a systematic way in their own work. Psychoanalysis, though early an established subspecialty in psychiatry, has enjoyed an uneasy relation with established medicine, as has psychiatry in general. In psychology, however, psychoanalysis received a generally sympathetic reception from psychologists representing diverse schools of thought, e.g., G. Stanley Hall, E. B. Holt, John B. Watson, and Robert S. Woodworth. The extent of general recognition of psychoanalysis was noted by D. G. Park (1931), who published in the *Psychological Review* an analysis of Freudian influence on academic psychology as measured by the extent to which psychoanalytic theory was included in 50 textbooks on general psychology. Park found that of the 50 books investigated, only four ignored psychoanalysis, and with the exception of one strongly anti-Freudian text by Dunlap, no text published after 1917 failed to show its influence. It is worth noting that John B. Watson's 1924 text reflected substantial psychoanalytical influence, as did the prominent 1929 textbook by Woodworth. Park notes that most of the texts agreed that Freud had overemphasized the place of the libido or sexual drive.

As Shakow and Rapaport (1964) properly emphasize, in spite of having favorable things to say about psychoanalysis and the increasing acceptance of aspects of Freud's conception of human nature, the attitude of academic psychologists must be described as mixed, since they made no

serious use of psychoanalytic theory in their own scientific work.² Shakow and Rapaport conclude that most academic psychologists were influenced by Freud primarily through the impact of psychoanalytic ideas on the Zeitgeist of the 20th century and thus were influenced in the same way as many other persons in the culture rather than as academic psychologists pursuing their own special studies; elsewhere Heidebreder (1933) has expressed a similar view. We think it fair to say that, from a scientific standpoint, psychoanalysis has had only a slight effect on the mainstream of academic psychology during the period covered by this review.

2. PSYCHOANALYTIC THEORY PERTINENT TO EDUCATION

As would be expected of a literature covering more than a half century and concerned with a theory of the complexity and subtlety of psychoanalysis, there are no succinct or simple formulations relevant to education we can refer to. Freudian theory has undergone at least four major revisions, and different writers at different times have concentrated on different aspects and interpretations of the theory. Though it is not possible within the confines of this article to give even a sketch of the history of psychoanalytic theory, there is one development that should be mentioned.

Freud's early model of the mind, the topographic model, marked a step away from the emphasis on consciousness shown by traditional philosophy and academic psychology. Though the model defined mental contents and processes on the basis of their availability to consciousness, it is clear that during the period of his early work, Freud's primary interest lay in that which remained out of awareness, that is, was unconscious. That the theory created a sensation because of the emphasis it placed on the power of the sexual drives on psychological functioning is by now familiar history and needs no comment here.

By 1923, various sorts of evidence had suggested to Freud that the attribute of consciousness was not adequate to differentiate between the regions of the mind, and in *The Ego and the Id* (1923/1961) he replaced the topographic model with his now famous structural model of id, ego, and superego. This second model added to the early interest in instinct a focus on ego function. Since that time, psychoanalytic investigators have continued to deepen their understanding of the various ego functions, so that psychoanalytic ego psychology is today concerned with functions that are not or only in a minimal way involved in conflict as well as with functions that mediate between the conflicting demands of id, superego, and envi-

² Notable exceptions to this generalization are E. R. Hilgard, H. A. Murray, and R. R. Sears.

ronment.³ Though different writers have different things to say about the nature and origins of these functions, ego psychology had become, by the end of the period of our review, an established part of psychoanalytic thinking. We mention this development because it puts in context the writings we discuss below.

In this section we review a representative sampling, through 1957, of the psychoanalytic literature relevant to ordinary elementary education. This literature can be regarded as important either because of the historical fact of concern about the issues discussed therein or because of the permanent intrinsic importance of these issues for education. The writings we consider concern learning theory, learning difficulties, moral development and discipline, and the teacher. The child's psychology from a psychoanalytic point of view is not discussed separately but is implicit in all the discussions.

2.1 *Anna Freud's Lectures to Teachers*

To give a concrete sense of the issues psychoanalysts consider to be of importance for education, we have selected what is perhaps the most famous single document on these matters during the historical period we are considering, namely, Anna Freud's lectures to Hort teachers in Vienna, which appeared in English translation as A. Freud (1935).⁴

The first two lectures concern the prelatency child, that is, the child before he enters school at the age of 5 or 6, and are devoted to the development of the following themes. First, one must consider education to begin on the first day of life and not when the child enters the formal school setting. Second, one must take the child's instinctual life into account if education is to succeed, or, to put the matter in another way, too narrow a view of the purpose of education is certain to fail. Education, whether administered by the parents or by the teacher, has as its general aim to make the child become like the adults around him, and to achieve this purpose the educator must engage in a constant struggle with the child's instincts.

Miss Freud lays out some of the main components of this instinctual life, emphasizing early sexual development, sibling rivalry, the Oedipus complex, or the male child's sexual love for his mother and rivalry with

³ Major publications in ego psychology include Anna Freud's *The Ego and the Mechanisms of Defense* (1936/1946), Heinz Hartmann's *The Ego and the Problem of Adaptation* (1939/1958), and Eric Erikson's *Childhood and Society* (1950).

⁴ A Hort as constituted in the Vienna of the 1930s may be compared to our present day-care centers, except that the children were older, ranging in age from 6 to 14 years. Thus, the teachers to whom the lectures were addressed were not ordinary schoolteachers but ones who related to the children in a more personal and less scholastic way than would ordinarily be the case. This fact, however, does not too much affect the content of the lectures, but only certain points of emphasis.

his father (she does not discuss the corresponding situation for the female child), and the transference of these family feelings to the school.

These particular features are all a natural part of the child's robust instinctual life, which unfolds through a series of stages. A vivid passage depicts this "natural state" which the parents or teachers must struggle to change:

The child is frightfully inconsiderate of others and egotistic; he is only concerned with getting his own way and satisfying his own desires; he is quite indifferent as to whether this hurts others or not. He is dirty and odoriferous; he does not mind catching hold of the most disgusting things or even putting them to his mouth. He is quite shameless so far as his own body is concerned and very curious about the things that other people wish to conceal from him. He is greedy and will steal dainties. He is cruel to all living creatures that are weaker than himself and filled with a perfect lust for destroying inanimate objects. He has an abundance of naughty bodily tricks, he sucks his fingers, he bites his nails, he picks his nose and plays with his sexual organs; he does all these things urged by his intense desire for self-fulfillment, and regards the slightest hindrance as intolerable. (pp. 46-47)

We have deliberately used the word *struggle* because one of Anna Freud's most important conceptual messages is that from the first day of life there is conflict between the aim of education, i.e., the transformation of the child into an adult of his particular society, and the instinctual life. The problem for education, as she emphasizes, is to be neither too repressive nor too permissive, and to understand that struggle and conflict are inevitable and should not be considered unnatural.

The development of the child has been hampered by the tradition of imposing education by force (as if it can be done), either by the threat of punishment or by the threat of the loss of love. The phenomenon of infantile amnesia is the residue of such measures taken to change the child, for as he adopts the viewpoint and values of the adults who are important to him he represses the memories of his early years, full of experiences and feelings no longer acceptable. The fact of infantile amnesia is important because it demonstrates how much of our own inner lives, including intimate knowledge of our past, is inaccessible to us even though it retains power to influence us.

The third lecture concerns the latency period, which is marked by the repression of the instincts characteristic of the earlier stages. Many mechanisms are at work during this period. A typical example is a reaction formation by the child toward his own excrement so that the desire to taste and touch it is replaced by a strong feeling of repulsion. Another example is the sublimation of an undesirable trait into a desirable one, so that, for example, an early preoccupation with urine and feces is transformed into pleasure in games with sand and water or with mud or paints. That the child is especially receptive to learning during the latency period has been recognized from time immemorial, and traditional

schooling is begun.

Adolescence, long regarded as marking the beginning of sexual feeling, in fact only witnesses a repetition of the earlier stages prior to the latency period. There is a revival of strong emotions due to the increased strength of the instincts, and the school must be prepared for this change.

The last of Anna Freud's lectures concerns the relations between psychoanalysis and pedagogy, which she summarizes as follows. There are three important clusters of ideas that psychoanalysis brings to our understanding of the child. The first concerns the division of his development into psychosexual stages. The second has to do with the growth of the child's personality, beginning with the instinctual life and followed by the development of the ego and the superego, which are derived from the influence of the parents and constitute the parents' primary contribution to his education. The third cluster of ideas concerns the interactions of the various parts of the child's personality. It is one of the significant mistakes of classical pedagogy to think that development can be a peaceful process and one of the most important lessons of psychoanalysis that the developmental process is inevitably a process of conflict.

Almost uniquely among important philosophies of education, psychoanalysis recognizes an explicit need to limit the goals of education. Without such limitations, education will tend to overly restrict the important part that should be played in life by the instinctual drives. In the psychoanalytic view, the profound problem of education is to find the right proportion of instinct gratification and instinct restriction. Although the psychoanalytic theory of pedagogy is not yet fully developed, it does bring three things to pedagogy: a basis for criticizing existing educational methods, a deepening of the teacher's understanding of the complicated psychological relations between the child and the teacher, and, finally, a practical therapeutic method for the treatment of children when psychological injuries have been "inflicted upon the child during the processes of education" (A. Freud, 1935, p. 106).

This brief description of the main themes of Anna Freud's lectures to teachers gives a summary of what psychoanalysis has to say about education. In the various subsections that follow we explore individual topics in more depth.

2.2 Theory of Learning

The discussions of learning in psychoanalysis center at least as much around the learning of the child in the early years of life as in the years of schooling; however, much of the discussion is general and applies to the later as well as to the earlier years of childhood. Broadly speaking, much of what is referred to as learning in the psychoanalytic literature is discussed under the heading of development in other parts of child psychology. For example, psychoanalysts label as a process of learning the

child's replacement of the pleasure principle by the reality principle, a step that requires him to reduce or delay immediate gratification and to suffer immediate deprivation in the interest of future gratification.

Because the psychoanalytic view toward learning emphasizes increasing differentiation of structure and capacity for sublimation, it is not surprising to find that a characteristic theme of considerable significance in psychoanalytic approaches to learning is the emphasis on readiness for learning or preconditions of learning. Psychoanalysts are undoubtedly right in saying that classical learning theories in academic psychology have not properly emphasized the deeper reasons for readiness for learning. The postulated stages of development, leading into the latency period around the age of 5 or 6 years, provide a conceptually compelling explanation of why this is the appropriate time to begin serious school learning—more so than is provided by any competing theory known to us despite the difficulties of formulating the theory in suitably precise form.

There are four related groups of articles that we would like to consider, even if somewhat cursorily. The first is much concerned with reward and punishment or what in learning-theoretic terms would often be called positive and negative reinforcement. In an influential article on the means of education, Fenichel (1945) summarizes matters this way: "And now it seems clear that we have identified the three basic means of all education, namely, direct threat, mobilization of the fear of losing love, and the promise of special rewards. The second means can be applied effectively only if the child has previously experienced the fear of losing love" (p. 285).

The fear of losing love is a characteristic psychoanalytic theme that is not ordinarily discussed as part of theories of learning but that has a powerful ring of truth about it and is in fact used in a variety of ways by parents and teachers as a means of directing the education of the young. Ideas of a similar nature, though not as systematically worked out, are expressed in Peller (1946). Peller has one viewpoint, however, that we think is fundamental to the psychoanalytic theory of rewards and punishments as a means of education. Here are her own words:

Threats and rewards are regulatory mechanisms effective in bringing about minor adjustments. The view that they can build the character structure which makes an adult out of a child, seems a replica of Lamarckian thinking. Lamarck explained the evolution of species by external, environmental regulation. . . . According to modern biology external regulatory influences could never lead to a new species. (p. 411)

Peller is expressing what has come to the fore in recent theories of learning, namely, the importance of postulating internal structures. The theories of learning she criticized certainly were not powerful enough to generate or account for the structures characteristic of the complex thinking and behavior of the child. It is characteristic of a psychoanalytic

viewpoint toward these matters that a balance is struck between structures that are inherent in the organism and the environmental forces that impinge from without. Although a variety of different opinions can be found among psychoanalytic writers about the details of these matters, there is a general insistence that neither internal structure nor environmental conditions alone can provide an adequate account of complex development. Along these lines, Hartmann, Kris, and Loewenstein (1946) propose four factors that are, in their view, the primary ones taken into account in psychoanalytic hypotheses about learning. We quote their brief summarizing passage: "First, the stage of maturation of the apparatuses; second, the reaction of the environment; third, the tolerance for deprivation; and fourth, the various types of gratification afforded by the processes of learning and the satisfactions that can be obtained as consequences of mastery" (p. 26).

The second group of papers on the relation of psychoanalysis and learning theory is concerned with what might be broadly classified under the heading of motivation. These articles deal with the specific roles of repression, sublimation, the developing dominance of the reality principle over the pleasure principle as the child enters school, and related phenomena. Excellent examples are the articles on sublimation by Hartmann (1955) and Peller (1956), a brief article on repression as a factor in learning theory (Harris, 1950), and an article on libidinal fixations by Liss (1935). It is not possible to explore these matters in detail here, but the point we do want to make is that there is a wealth of additional psychoanalytic concepts that has not received much attention in the mainstream of learning-theory research in this century.

A good summary of the situation is provided in Hilgard (1948/1956). Hilgard provides an especially good account of attempts to experimentally test various psychoanalytic concepts and propositions. Though it is outside the scope of this chapter to review this literature, the reader who wants to gain a sense of these matters can certainly do no better than consult the excellent survey and detailed list of references given by Hilgard. He provides a good account of psychoanalytic concepts that have influenced learning theories and experiments as well as an excellent précis of those psychoanalytic ideas that had had, as of 1956, little impact on psychological studies of learning.

There is one point about Hilgard's survey and the literature he references that is of some general interest. It might have been thought that Freud's concept of psychic energy would have been taken up by psychologists, especially learning theorists, and developed into a precise and quantitative concept, testable against a variety of empirical phenomena. Somewhat surprisingly, this has not been done, and in fact we know of no theoretical discussion of the concept of energy by a learning theorist or of any attempt to provide an empirical test of the

concept. We also know of no discussion of learning by a psychoanalyst that brings the concept of energy and the related important concept of cathexis to bear on theoretical problems of learning, in spite of some tantalizing remarks here and there in the psychoanalytic literature on these matters and the obvious importance of these ideas to any development of learning theory along psychoanalytic lines.

The third group of articles, of which we again give two sample instances, is concerned with the relation between psychoanalysis, conditioned reflexes, and stimulus-response psychology. We consider here the articles of Bernstein (1935) and Schilder (1937). Bernstein prefers to contrast psychoanalytic approaches to learning with Thorndike's basic laws of learning rather than to sharply oppose them. For example, he says: "The pedagogue is seeking *positive* responses and attempting to reduce the number of *negative* responses. Psychoanalysis makes a more complicated analysis of interest and works back to the intricate concept of impulse-disposition and to the unending conflict between the ego and the id. The analyst emphasizes the divergences from the desirable pattern while the statistically-minded educator would stress the successful responses and draw conclusions therefrom" (p. 159). Bernstein also notes that from a psychoanalytic standpoint there is a sense in which all education represents a form of repression.

Schilder's article is deeper in its analysis of the relations between psychoanalysis and stimulus-response theory, especially as reflected in Pavlov's theory of conditioned reflexes. He is critical of the acceptance of conditioned reflexes as adequate to explain complex human behavior, and is especially telling in his detailed critical analysis of the use of physiological language to describe what are, in fact, psychological phenomena. Some of his remarks are amusing as well as to the point; for example, "The neurosis in dogs as described by Pavlov certainly has a primitive structure. It can be cured by rest, bromides and calcium and if there is anything by which a neurosis in humans cannot be cured it is rest, bromides and calcium" (pp. 13-14).

Schilder writes from a medical as well as a psychological background. The following passage, critical of Pavlov's theories of inhibition and excitation and of other simple attempts at localization of psychological processes in the nervous system, provides a good sense of the specific character of Schilder's criticisms.

Of course, nobody has a right to doubt that psychological processes are dependent from cortical processes. Still I doubt the formulations concerning excitation, inhibition and irradiation. The interesting experiments concerning the corpus callosum can be explained in the following way from a psychological point of view. The knowledge of one's own body has to be built up. In this process of building up, the impressions of both sides of the body have to be continually coordinated. There is no question that the process of coordination

will be disturbed by transection of the corpus callosum. But, I do not believe that we deal merely with symmetrical connections of excited points. (p. 16)

The last group of articles we review focuses on the topic of curriculum and instruction. There are relatively few psychoanalytic writings on these matters during the period we review, and because these articles concern the effects on learning of various approaches to the curriculum, it seemed appropriate to include a discussion of this issue in this section.

Zachry (1933) discusses the role of instruction in art for personality development and presents a case for the inclusion of art in the curriculum. Bettelheim (1948) discusses the ways in which the social studies teacher can help students master the problems of adolescence. He emphasizes that real-world mastery of his problems is not possible for the adolescent so that classroom analysis is especially important.

An example of a detailed discussion of particular parts of the curriculum from a psychoanalytic standpoint is to be found in Hill (1971). Although this book appeared after 1957, Hill has published extensively on psychoanalysis and education, beginning with a paper "Poetry and the Unconscious" in 1924, in which he expresses views similar to those expressed in his book. A shorter discussion, also dated after 1957, on how psychoanalysis throws light on the formulation of educational objectives is to be found in Tyler (1958).

2.3 Learning Difficulties

The clinical viewpoint of psychoanalysis has been applied to a wide variety of learning difficulties encountered by students, and the results are of considerable conceptual interest. The literature on this subject is large, so, as with other topics, we present a sample of the kind of thinking that has been typical.

In the early psychoanalytic view, learning inhibitions result from unstable sublimations of instinctual drives. This idea may be traced to Freud (1926/1959), who wrote:

The ego function of an organ is impaired if its erotogenicity—its sexual significance—is increased. . . . As soon as writing, which entails making a liquid flow out of a tube on to a piece of white paper, assumes the significance of copulation, or as soon as walking becomes a symbolic substitute for treading upon the body of mother earth, both writing and walking are stopped because they represent performance of a forbidden sexual act. The ego renounces these functions, which are within its sphere, in order not to have to undertake fresh measures of repression—in order to avoid a conflict with the id.

There are clearly also inhibitions which serve the purpose of self-punishment. This is often the case in inhibitions of professional activities. The ego is not allowed to carry on those activities, because they would bring success and gain, and these are things which the severe super-ego has forbidden. So the ego gives them up too, in order to avoid coming into conflict with the super-ego.

The more generalized inhibitions of the ego obey a different mechanism of a

simple kind. When the ego is involved in a particularly difficult psychical task, as occurs in mourning, or when there is some tremendous suppression of affect or when a continual flood of sexual phantasies has to be kept down, it loses so much of the energy at its disposal that it has to cut down the expenditure of it at many points at once. (pp. 89-90)

The most conspicuous manifestation of intellectual inhibition during the early years of school life is in reading difficulty. There is a large psychoanalytic literature on this topic, beginning in the 1920s, a time when the general psychological literature on reading also began to flourish. An excellent systematic survey of psychoanalytic contributions to the problem of reading disabilities is to be found in Blanchard (1946), who presents a survey of the statistical data on sources of reading difficulty and, referring to Gates' extensive study, accepts his estimate that emotional disturbances are present in about 75% of children with serious reading disabilities, that in about a fourth of these cases the emotional difficulties are the cause of the reading difficulty, and that in the remaining three-fourths of the cases the causal direction is reversed. She emphasizes that psychoanalytic therapy is especially called for in cases where neurotic difficulties are the source of the reading disability, whereas therapy may not be required when the causal direction is reversed, except in severe cases.

Blanchard subscribes to the theoretical view that learning offers an opportunity for the sublimation of instinctual drives and explains reading disabilities this way:

More frequently the child will be suffering from severe unconscious conflicts, with repression of impulses and imperfect capacity to sublimate them, so that much of his energy is used up in maintaining repression and not enough is left over for such a complex mental process as learning to read. Moreover, this type of child often tends to resort to restriction of ego activities in order to escape painful situations and now may utilize this defense mechanism to avoid the painful experience of seeing classmates excel his achievement, giving up all effort to learn to read as a way of evading competition where he knows he will appear inferior to others. (p. 170)

She also points out the role played by chronic emotional stress in the child's background in producing disabilities. This stress does not necessarily result only in reading failure, but since reading failure is the main source of general school failure in the early years, in a more pervasive kind of failure as well.

A good survey of the literature on emotional components of learning arithmetic is given in Plank and Plank (1954). Their method of study consists in the examination of autobiographies for remarks about the pleasure or pain of learning mathematics. Though their methodology is flawed, their results support the claims of several writers, e.g., Levy (1943) and Klein (1949), that overprotected children who are strongly attached

to their mothers often find learning to read easy and learning arithmetic difficult.

There are other instances in which psychoanalysts have brought a clinical-case approach to bear on the issue of how affective factors may interfere with cognitive processes. Klein (1924) gives a rich sample of such clinical analyses.⁵ She brings out in the context of detailed case studies some of the fantasies that may stand in the way of proper learning and that are in a poorly understood fashion part of ordinary folklore. Fantasies about particular letters and particular numbers are perhaps the most familiar examples.

For little Fritz in *writing* the lines mean roads and the letters ride on motor-bicycles—on the pen—upon them. For instance, 'i' and 'e' ride together on a motor-bicycle that is usually driven by the 'i' and they love one another with a tenderness quite unknown in the real world.

Thus, instead of the double 's', he always wrote only one, until a phantasy afforded the explanation and solution of this inhibition. The one 's' was himself, the other his father. They were to embark together on a motor boat, for the pen was also a boat, the copy-book a lake. The 's' that was himself got into the boat that belonged to the other 's' and sailed away in it quickly upon the lake. This was the reason why he did not write the two 's's' together. (p. 318)

This kind of analysis extends to even so specific and technical a matter as the algorithm of division. Here is a statement about Fritz's problems:

Fritz had a marked inhibition for doing division sums, all explanations proving unavailing, for he understood them quite well, but always did his examples wrong. He told me once that in doing division he had first of all to bring down the figure that he still required, he climbed up, seized it by the arm and pulled it down. To my enquiry as to what it said to that, he replied, that quite certainly it was not pleasant for the number—it was as if his mother stood on a stone 13 yards high and someone came and caught her by the arm so that they tore it out and divided her. Shortly beforehand, however, he had phantasied about a woman in the circus who is sawn in pieces, and then nevertheless comes to life again, and now he asked me whether this were possible. He then related (also in connection with a previously elaborated phantasy) that really every child wants to have a bit of his mother, who is to be cut in four pieces; he depicted quite exactly how she screamed and had paper stuffed in her mouth so that she could not scream, and what kind of faces she made, etc. (p. 324)

That these fantasies are not unique to early learning but may interfere with later learning is shown by the following case:

Seventeen-year-old Lisa related in her associations that she did not like the letter 'i', it was a silly jumping boy who always laughed, who was not needed in the world at all and over whom she became quite enraged, incomprehensibly to herself. She praised the letter 'a' as being serious and dignified, it impressed

⁵ Although Melanie Klein is not considered an orthodox psychoanalyst, we present this material here because of its vivid clinical details.

her and the associations led to a clear father-*imago* whose name also began with an 'a'. (p. 319)

Here are further remarks about Lisa's difficulties with arithmetic:

Lisa considered the number '3' also insupportable because 'a third is of course always superfluous' and 'two can run races with one another'—the goal being a flag—but the third has no business there. Lisa, who had a taste for mathematics, but was very inhibited where it was concerned, told me that really she only thoroughly understood the idea of *addition*; she could grasp 'that "I" joins with another when both are the same', but how were they added up when they were different? This idea was conditioned by her castration complex, it concerned the difference between the male and female genitals. (p. 321)

From the standpoint of the academic literature on students' performance, these clinical accounts may seem fanciful. It is not our purpose here to defend in detail the correctness of Klein's account but rather to emphasize that she has undoubtedly identified some of the reasons for individual children's difficulties. Most of her examples have to do with reading and arithmetic because of their central role in school performance but she also includes examples from grammar, geography, and physics.

The articles cited above discuss instances in which the child's ego is overwhelmed by other forces. In the last article we review, that by Pearson (1952), we find a discussion of learning disabilities understood as a problem in ego psychology as such. Pearson discusses at some length the child who seems to lack any interest in learning, and for whom learning is boring because it requires effort and interferes with immediate pleasures.

Pearson is raising an issue that has increasingly engaged the attention of educators and psychoanalysts, that of the child whose ego development has been retarded by a too permissive upbringing. We discuss below how this kind of childrearing philosophy has been one of the most persistent misapplications of Freud.

2.4 *Discipline and Moral Development*

The disciplining and moral development of children have been traditional concerns of those responsible for formulating educational aims and for organizing society's efforts at schooling. The general history of the subject extends back at least to ancient Greek philosophy where it is discussed extensively in various works and from various viewpoints by Plato and Aristotle.

To a large extent, the psychoanalytic view of learning and moral development is related to the theory of learning outlined above, since psychoanalytic writings on these topics are more concerned with the promotion of inner control than with overt behavior. The term "discipline" is perhaps best used with younger children, since the young child does not have the inner resources for regulating his behavior; moral

development, on the other hand, is concerned with the acquisition of internal structures that permit the child to regulate his behavior even when his parents, or parental surrogates, are absent.

There is in psychoanalysis a rather distinctive theory of moral development. This theory, especially as it relates to the theory of the superego, was not articulated in finished form in Freud's early writings, but was extensively developed over the period covered by this review. By the end of the period covered by our review, the psychoanalytic position was approximately as follows. The superego develops fairly early in childhood, certainly by the age of 5 to 8 years. The development of the superego is not to be identified with the development of moral conscience, but rather moral development is an important part of superego development. The mechanism by which the superego develops is the child's identification with his parents and with parental substitutes such as teachers. Later writers, e.g., Hartmann and Loewenstein (1962), have argued strongly for the position that development of the superego continues through adolescence and represents a continual adjustment with ego functions as well as with the entire social and cultural framework in which the young person is maturing.

In general, the psychoanalytic position on discipline and moral development has been misunderstood. A common popular image is that psychoanalysis has encouraged, as has no other movement, the undisciplined child who is permitted to develop in his own way and without the external constraint that he conform to the attitudes and habits of adults around him. This widely held popular view is a caricature of what psychoanalysts actually write about the problem of discipline. Indeed psychoanalytic writings on this subject stress the need for avoiding any simple approach to these problems. On the one hand, impulses must be allowed to surface and to be expressed; on the other hand, restraints on his instinctual demands must be imposed or the child's ego development will suffer. Put somewhat more explicitly and theoretically, the central role of discipline in child development is to support the ego in its attempts to gain control over the id, or instinctual impulses.

These ideas are spelled out with particular clarity in a classic article by Sperling (1951) and are also discussed with sophistication and perception in Kris (1948, 1950). Sperling is explicitly and properly critical of the attitudes toward discipline promulgated by A. S. Neill and made famous by the practices at Summerhill. Kris (1948) emphasizes the misunderstandings that arose concerning the use of indulgence and deprivation as means of education. In the greater part of his article he discusses the importance of establishing a dynamic balance between indulgence as expressed in parental love, or the teacher's tolerance of unruliness, and deprivation, meant to include all expressions of the educator's disapproval, from the denial of a smile to disciplinary methods of greater

severity. Though Kris emphasizes the importance of indulgence, he does not let stand the view that the application of psychoanalytic theory to education leads to the elimination of deprivation or discipline. An excellent general statement by him on these matters is the following:

In discussing the dangers of a too permissive attitude one may seem to advocate returning to an outdated method of dominance in handling the child. Nothing could be less desirable. We are, to quote Freud, dealing with the avoidance of two dangers, of Scylla and Charybdis. While indulgence and deprivation may both create unwanted effects, both are essential measures of any education. Both meet with some of the child's needs; the question is one of modality and timing. (p. 626)

Kris also emphasizes the importance of deprivation or discipline for providing ego support; he spends a good part of the article in discussing the structural model and the importance of ego development in education.

2.5 *The Teacher*

One of the more insightful discussions of the teacher from a psychoanalytic standpoint is to be found in Bernfeld (1925/1973), who was one of the prominent psychoanalytic pedagogues in Vienna just after World War I. Bernfeld points out the psychological dangers inherent in the teaching situation and laments the ordinary teacher's too frequent lack of insight into this problem. Unfortunately, Bernfeld's monograph on the limits of education was not translated from the German until 1973 and thus cannot be said to have had a major influence on American thinking during the period under review, although it is referred to in the original in a number of the publications in American journals.

A good statement of how the position of teacher can give rise to many problems is to be found in Low (1938):

One of the first importance is due to the scope and outlet provided for aggressive elements in the personality, elements wholly unsuspected, perhaps, by their owner. Analytic investigation has revealed that aggression is a powerful factor in every human being's make-up, expressing itself positively and negatively, but often most deeply-concealed and most heavily disguised, and on that account ready to seize all available outlets. The authoritative position of the teachers, which the majority of the taught so readily support—owing to their unconscious love and fear motivations—gives sanction to the exercise of aggression in the name of 'Discipline', 'Training', 'Standards' and such a hidden trend may be found existing side by side with a wholly different conscious attitude.

And this leads me to the consideration of what is perhaps the most difficult and yet most inevitable problem of the teacher—the emotional relationship between teacher and taught. Quite a large proportion of the teaching world, in unconscious fear of facing this relationship, tries to 'solve' the problem by denial of its existence. . . . If the teacher's own emotional needs are still very little fulfilled outside the school-world, and too strong to be ignored, it may be

impossible to dam up the stream, with the result that it is unduly poured out on the pupil and a too close, fictitious bond may be created between the two, handicapping the latter's development. In extreme forms—those dealt with by novelists and dramatists—this situation is easy to recognize (though by no means always easy to deal with successfully) but the more subtly disguised and less wholesale ones may pass almost unrealized. (pp. 228–229)

Other authors address the problem of how to differentiate the role of the teacher from the role of the therapist. A British publication by Searl (1932) draws a number of useful contrasts between psychoanalysis and education.

Searl emphasizes the usefulness of a personal psychoanalysis for teachers but at the same time expresses the limitations of what might be expected, for, as he puts it, "no amount of analysis can turn a really unsuitable person into a good pedagogue" (p. 290). His views on these matters are especially of interest because he explicitly emphasizes that he is summarizing the views of a group of analysts and educators on the value of psychoanalysis to educators. Searl emphasizes the problems and, in many cases, the dangers of partial knowledge of psychoanalytic theory. An example he gives is the following:

This is the case of an intelligent, able and sympathetic head mistress, who had read a good deal of psychoanalytic literature and had become very impressed with the idea that "repression" was to be avoided at all costs. When a little girl died in her school, the head mistress wrote to the mother inviting the small sister of the dead child to come and stay in the school for a time, and talk about her loss to the people who had known her sister, on the ground that she "must not be allowed to repress her grief." Now this again was a woman who would not have dreamt of doing such a thing out of her ordinary standards of what was kind and wise. Her normal judgment was distorted and her understanding blurred by the tyranny of a theory. (p. 291)

Searl, however, does indicate positive ways in which a knowledge of psychoanalytic theory can be helpful to the educator. One is in helping the educator to set standards of what is to be regarded as normal behavior in a child of a given age. A second and related source of assistance is in enabling a teacher to see when a child is genuinely emotionally disturbed. A third is in offering a deeper understanding of educational method, for example, the importance of play for the child's development. Nevertheless, Searl emphasizes the importance of separating the roles of educator and analyst.

An article appearing at about the same time, much referred to later, is that by Erikson (then Homburger; 1935). In the opening passages of this article he describes very well the contrast between the role of the teacher and the role of the therapist.

Of all those who through their analytic training hope to be able to make some fundamental contribution to therapeutic or educational work, the teacher is the least able to foresee what he may achieve through analytical insight, which

he gains from his own clinical analysis. The analytic situation does not offer him any direct suggestion as to how to face the specific situations he meets on returning to his work. An analyst is obliged for the most part to remain a silent observer while the teacher's work involves continuous talking—this fact alone roughly distinguishes the methods of analyst from that of teacher, representing the two extremes of all possible educational methods of approach. (p. 50)

Erikson brings out through some detailed examples the dilemmas teachers face when answering children's questions, for example, the difficulty of knowing how to respond to what one sees as the hidden meaning, often sexual in nature, of the question asked. Here is an unusually perceptive passage on these matters that is specifically addressed to the problem of sex education:

That the adult who is questioned by a child is in the position to give interpretations seems to be only a first step forward. Above all, the adult must know that consistent and effective interpretation belong in the realm of clinical analysis. Then, once he recognizes the twofold meaning of the child's questions he is faced with two possibilities. He may either ignore the hidden meaning and consequently answer inadequately—perhaps even more dangerously and less adequately than the stork tale—or, disregarding the disguise of the question, he may interpret its hidden meaning, answering more than was intentionally asked. This may prove a shock for the child or, as is more probable, it will remain absolutely ineffective. And few things undermine the position of the adult more disastrously than serious but ineffective effort!

For the teacher there remains another individual problem. He must not only appreciate the sexual curiosity masquerading as desire for knowledge, but he must make the greatest possible use of it. The child never learns more than he does at the time of disguised curiosity. At this time he learns with the cooperation of his affects, and now when he hopes finally to find out "the hidden secrets," the statement "*vitae non scholae discimus*" really holds good, for he is learning now for the sake of the life he dimly divines, and not for the sake of his lessons.

One might think that the teacher could make use of this situation to smuggle into his answer the sexual enlightenment that the child's questions have unconsciously requested. By general frankness he should be able to establish the certainty that there is nothing more secret about sex than about everything else. However, glimpses into the unconscious, such as described above, show that the sexual questions as they reach a complicated and dangerous point (a moment which enlightenment is supposed to guide) inhibit the desire for further questioning and the wish to learn. Softly the child speaks of the purple snail and fire engines, the food tube of the hen and the goiter that some women have. And still more softly come the inmost questions of the child which only barely make themselves heard to his interpreter: "What about the desire and the fear of destroying, and the fear and desire of being destroyed?" With his most earnest questions, then, the child still remains alone. (pp. 62-63)

We have cited this lengthy passage because it puts in vivid terms what is perhaps the most important and yet extraordinarily subtle message psychoanalysis offers the teacher. Characteristically, there is no simple approach to the problems of education. Without offering any panacea for

these complicated problems, Erikson concludes on the optimistic note that self-knowledge cannot help but make for more effective teaching and that a personal psychoanalysis for the teacher should provide a powerful means for its achievement.

In an article, written toward the end of the period we review, Anna Freud (1952/1968) discusses three pitfalls the schoolteacher must beware of. They are these. First, that it is unwise to work with only one age group, a practice that encourages the teacher to lose sight of the fact that any specific age is merely one stage within childhood. Second, that too close an alliance with young children to the exclusion of adult contacts fosters a loss of perspective. Third, that too strong an attachment to an individual child must bring with it inevitable frustration and conflict with the child's mother. Miss Freud suggests that a useful way out of the dilemmas faced by the teacher is to take a more general interest in the whole process of development.

In this article Miss Freud responds directly to teachers' questions, and her responses underscore her view that the role of the teacher is distinct and different from that of the therapist. We quote some of the teacher's questions, as well as Anna Freud's responses, below.

If we adopt this point of view, don't we become impersonal in our relations with the children?

If the teacher sees in a child something that reminds her of her own childhood, she wants to help that child and save him from what she experienced. . . . But what of a child who reminds her either of a brother with whom she fought constantly or of something in her own development which she cannot face? For every child who benefits because he means something in her life, another child suffers. This is the danger of the personal approach. . . . A teacher cannot be completely objective, but once she becomes interested in the processes of childhood, all children become more interesting in a more objective way.

Would you discuss the problem of the child who comes from a home in which he is emotionally deprived?

To satisfy a deprived child's need for affection the teacher may want to play the part of the mother. She had better know that in this case the child will cease to accept her in the role of teacher. The teacher's role is not that of a mother substitute. If, as teachers, we play the part of a mother, we get from the child the reactions which are appropriate for the mother-child relationship—the demand for exclusive attention and affection, the wish to get rid of all the other children in the classroom. This makes it more difficult to deal with the child. Although it fills a need in the child, it is not the teacher's role.

It is inevitable that in the classroom the children will act out the attitudes which they transfer from home. But, normally, when the teacher stays within her role, the home situation does not develop fully in school. It is the privilege of the teacher to introduce the child to a new experience, to life within a social community, not merely to duplicate his experiences within the family.

Should the teacher try to help the problem child?

It may depend on the child's age. Helping the child with personal problems by understanding them is within the teacher's role. But if this presupposes as close a relationship between the child and the teacher as between a child and a therapist, then it transgresses the limits within which the teacher can work. She can help the child by varying her demands on him and her behavior toward him. But it should be remembered that the teacher is neither a mother nor a therapist. (pp. 562-565)

These articles attempt to tackle the problem of the means by which the educator may benefit from the findings of psychoanalysts without thinking of himself as a child analyst. Nevertheless, the literature we have cited refers chiefly to problems of conflict as they are encountered in the classroom and does not relate psychoanalytic insights to problems of education in which there is a more conflict-free didactic task.

3. EVIDENCE OF IMPACT

The assessment of the impact of any new concept or theory on education is necessarily a complex matter. From one viewpoint the ideal evidence in the present instance would be data obtained from extensive observations of elementary-school classrooms during the period 1900-1957, but even such data would not eliminate difficult problems of analysis. In any case, such data are not available, and our approach to assessing the impact of psychoanalysis on elementary education must be indirect. Within a limited framework of resources, we have restricted ourselves mainly to the published educational literature, including some secondary sources on courses and other institutional aspects of teacher training.

There has not been a psychoanalytic movement in American education, but other movements have, to a greater or lesser extent, carried psychoanalytic ideas into the schools. Perhaps the earliest of the movements that related to psychoanalysis was the child-study movement, often considered to have begun in 1880 when G. Stanley Hall returned from Europe and began his famous study on the contents of children's minds. An offshoot that might be claimed for the child-study movement is the science of child development, including child psychology, which became an established part of academic psychology in American universities during the 1920s. We discuss the child-study and child-development movements in Section 3.1.

The end of the 19th century also saw the beginning of the educational psychology movement, and the early applications of scientific methods to the study of educational problems. The educational psychology course quickly became one of the foundation courses in American teacher-training institutions, and in Section 3.2 we examine aspects of the educa-

tional psychology movement relevant to our investigation.

From a different direction and for different reasons, there arose sometime toward 1908 an organized interest in bettering the conditions of life for the mentally ill and mentally defective. Early efforts along these lines developed into what is known as the mental hygiene movement, to which substantial credit must go for changing the character of American education. In Section 3.3 we examine the impact of psychoanalysis on this movement.

All these developments were to a considerable extent part of the progressive education movement, which transformed American education over a period of at least half a century. Progressive and psychoanalytic thinking have often been confused, and it is of interest to discuss in a general way the extent of psychoanalytic influence on the progressive education movement. We do so in Section 3.4.

The literature on which psychoanalytic ideas might be expected to have had some impact, even within a given tradition, naturally falls into several different, though overlapping, categories. There seem to be at least three classes that are easily identified. One consists of textbooks written primarily for students planning to become teachers. To what extent the students absorbed the contents of the textbooks and, if they absorbed them, to what extent they put them to use in their own classrooms are not questions we can answer. Nonetheless, analysis of the extent to which psychoanalytic ideas are to be found in textbooks provides at least circumstantial evidence of influence.

Another class of literature is made up of educational journals that are written to be read by teachers or at least by selected groups of practitioners. The range of publications in this category was beyond our capacity to examine; there are, for example, many "local" journals indigenous to a particular state or region in this category. Because of the innovative importance of the progressive education movement during the first half of this century, we have focused our analysis on *Progressive Education*, the main journal of this movement, along with the closely associated *Teachers' College Record*. Our justification for selecting these journals is based on the proposition that it is of interest and significance to examine the trends in an innovative movement that had great overall importance for American education. Assessing the literature of progressive education is a partial surrogate for assessing the ideas of the leaders among educational practitioners.

A third category of literature that we briefly survey consists of an extensive miscellany of magazine articles and other publications written for the general public, and can be considered an indirect source of impact on education since it came from the surrounding culture. The brevity of our survey presented in Section 3.5 belies the importance we attach to this literature, for it is possible that the discussion of psychoanalytic ideas in

general publications had a greater impact on teachers than what was said in all the professional sources combined.

3.1 *The Child-study and Child-development Movements*

History of the movements. G. Stanley Hall was a vigorous, active, and opinionated man. In 1887 he founded the *Journal of Psychology*, the first psychological journal in the United States. A few years later he was instrumental in founding the Department of Pedagogy at Clark University, as well as the *Pedagogical Seminary*, the first official publication dedicated to child study. All during this time he engaged in a program of many public lectures around the country, and under his driving leadership the child-study movement spread rapidly to other universities and occupied a leading position among educational movements during the first decade of this century. A summary of Hall's activities and of the subsequent years of the child-study movement is to be found in an unpublished dissertation by Dutton (1945).

Many have remarked on G. Stanley Hall's interest in psychoanalysis. He probably wrote as much about Freud during the first decade of the century as did any of the physicians interested in psychoanalysis; in 1908 as president of Clark, he was responsible for bringing Freud and Jung to the 20th-anniversary celebration of the founding of the university. Because he was the acknowledged leader of the child-study movement, Hall was certainly influential in bringing psychoanalytic ideas to this country.⁶ However, his own works were not developed with a psychoanalytic view explicitly in mind, but rather Hall seems to have been an early representative of the eclecticism that was to characterize most of the child-development literature to follow; at his convenience he used psychoanalytic concepts to make certain points. His chapter on the pedagogy of sex, to be found in his two-volume treatise *Educational Problems* (1911), for example, cannot be said to be organized within a psychoanalytic framework, although it does mention Freud and is in other ways forward-looking for its day.⁷

⁶ Probably the first published American dissertation in psychology on psychoanalysis was written under Hall's direction, namely, Chase (1910).

⁷ A much later review (Low, 1923) of Hall's *Aspects of Child Life and Education* (1921) puts very well and in greater detail the point we are making:

From the psycho-analytical standpoint, this book provides excellent material, but it can hardly be said that the various writers have treated their material from such a point of view. In two of the most interesting essays—'A study of dolls' and 'Curiosity and interest'—it is striking that, in spite of ample material for the purpose, references to and conclusions concerning the unconscious mind are almost absent. One feels that the absence of psycho-analytic knowledge goes a good way towards vitiating the work. For instance, in the essay on 'Curiosity and interest', the meaning of curiosity, its sexual source and its emotional motivation seem quite unrecognised, and the distinction made between what is termed 'legitimate' and 'perverted' curiosity is not truly based. In the section dealing with curiosity relating to the origin of life, a good deal of misunderstanding is apparent: the authors appear to think that intellectual knowledge, correct information given at appro-

The child-study movement was for the most part self-stimulating in its development; in contrast, the science of child development was born of converging pressures to generally better the child's lot. One tradition was service-oriented and organized around case studies. The importance of this tradition, as reflected in the mental hygiene movement and the growth of child guidance clinics beginning with William Healy's first efforts in Chicago in 1909 (Healy, 1915), has been in the richness of clinical experience it has brought to the child-development movement. Later developments in the child guidance clinic movement include the organization of the American Orthopsychiatric Association in 1924 and the establishment of the American Association of Psychiatric Clinics for Children in 1948. The history of the mental hygiene movement is discussed in Section 3.3.

The second tradition was oriented toward the laboratory and interested in becoming respectable academic science. An early step in this direction came in 1906 when Mrs. Cora Bussey Hillis conceived the idea for what was to become the Iowa Child Welfare Research Station. Later, in 1922-1923, the first official step in establishing the science of child development was taken with the formation of the subcommittee on child development under the auspices of the Division of Anthropology and Psychology of the National Research Council. Readers interested in the details of the founding of the movement will find an excellent history and analysis in Sears (1975); earlier historical reviews that also discuss the place of psychoanalysis are Jersild (1948-1949) and Anderson (1956).

Textbooks. One measure of the impact of a theory on a profession is the extent to which it is represented in the standard textbooks. For this reason, we looked at all the more frequently cited texts in child study and child development used in the training of teachers for the years 1903-1956.⁸ Figure 1 shows the results of our analysis of 36 textbooks, which are listed below in Table 3. Each text has been read to determine the number of pages giving either explicit, implicit, or derived coverage to psychoanalysis. *Explicit* coverage refers to pages on which either Freud, psychoanalysis, or some other orthodox psychoanalyst is mentioned; *implicit* coverage to pages not included in explicit coverage but on which psychoanalytic concepts like that of the defense mechanisms are mentioned; and *derived* coverage to pages that mention neo-Freudians like Karen Horney or Harry Stack Sullivan, projective tests, and experimental attempts to test psychoanalytic ideas as well as attempts to combine psychoanalytic with other approaches (e.g., Dollard & Miller, 1950).

prate stages, will solve this difficult problem, thus ignoring the whole matter of the fantasies of the unconscious mind which are so profound an influence. (p. 211)

⁸ We selected these texts by examining the much larger number available in the Stanford University Library system and we believe that the texts selected are more or less representative. Selection of the texts reported in later sections was made on the same basis.

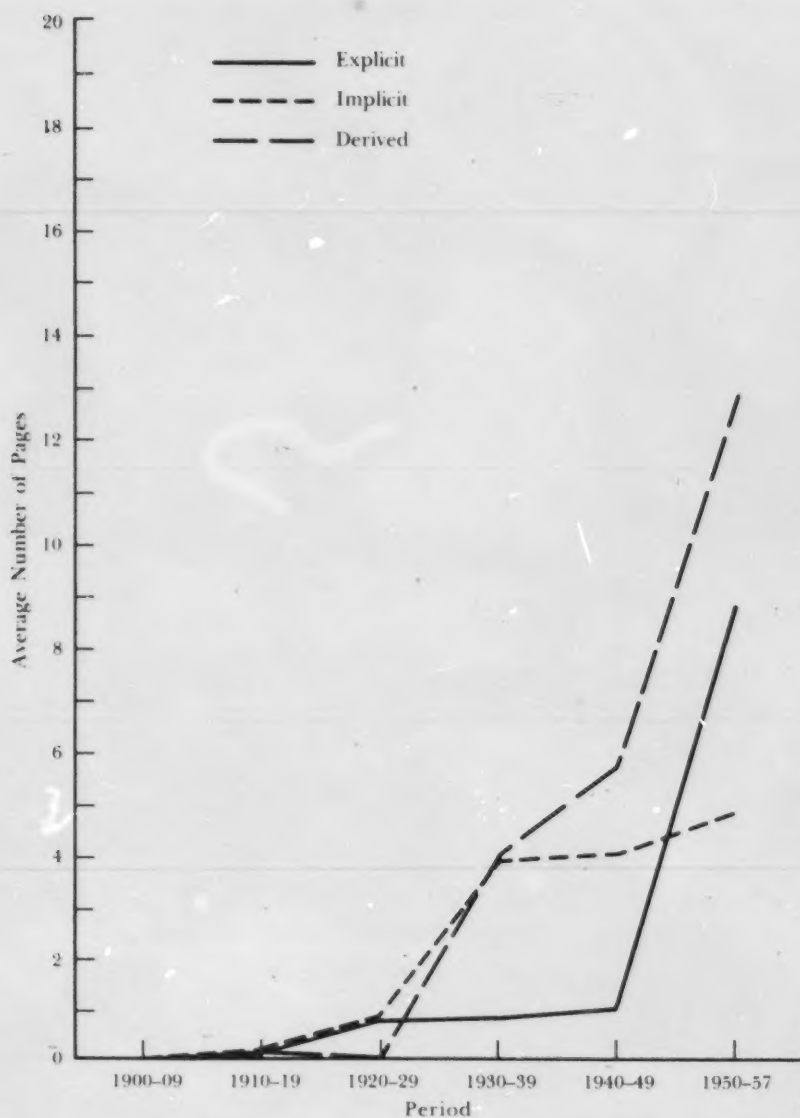


FIGURE 1. Evidence of psychoanalytic influence in child-study and child-development textbooks.

The data for Figure 1 were obtained by determining average pages for each decade. The increase in all three curves over time shows the increasing influence of psychoanalysis, even though the average number of pages remains relatively small. As is evident, substantial increases in coverage of explicit and derived topics occurred after 1940, and some of this effect was undoubtedly due to the influx of European psychoanalysts in the 1930s. The earlier and subsequently greater increase in the coverage of derived topics probably reflects the closer relation between projective tests and similar topics, on the one hand, and the primarily behavioral orientation of child development during this period, on the other. Although we fully recognize that the kind of analysis made is partially subjective and subject to error, we believe the overall trends shown in Figure 1 correctly reflect what in fact occurred.

Table 1 summarizes the explicit coverage of psychoanalytic topics in the texts and highlights several changes. For one, we note the absence of discussion of infantile sexuality during the early years. Indeed, explicit discussions of psychoanalysis during the years before World War II consisted exclusively of remarks about the psychoanalytic method, the importance of the early years for later development, and the Freudian theory of forgetting. After World War II, however, we find that 6 of the 10 books listed in the table discuss infantile sexuality, although for reasons we do not understand, even after 1940, no attention was given to the theory of aggression.

The second change of note concerns the increased spread of coverage, as reflected in the wider range of topics covered in the later years, as shown in Table 1.

The limited representation of psychoanalysis in the literature can be partly explained by the fact that all the texts cover an enormous amount of material and a wide range of topics. It has been characteristic of the movement to draw on work done by a wide spectrum of scientists, and much of each of the texts is devoted to material that is nonpsychological in nature, interest in child psychology as such being largely a recent development. An analysis of the topics covered, including the percentage of the total space given to each of the topics, is shown in Table 2. The data were averaged for each decade (with the percentages shown rounded to whole numbers), and they clearly show the diversity of topics and wide range of emphases. The table also points to a decreasing interest in biological functions and an increasing interest in social development and personality over time. On the other hand, coverage of learning, the emotions, individual differences, moral development, mental hygiene, and motivation remains relatively constant through the decades and the topics receive little attention.

Also characteristic of the texts is the fact that throughout most of the period covered by our review, the authors are highly restrained about

TABLE I
 Psychoanalytic Topics Given Explicit Coverage
 in Child Study and Child Development Texts

Author	Year	Topics													
		H	Me	I	F	L	P	Mo	C	D	Dr	Pl	Ch	In	En
Kirkpatrick	1911	.	.	.	.	.	.	.	X	.	.	.	.	.	.
Tanner	1915	.	.	X	X	X	.	.	.	.	.	.	.	.	.
Norsworthy & Whitley	1918	.	.	.	X	.	.	.	.	.	.	.	.	.	.
Waddell	1918	.	X	X	.	.	.	.	.	.	.	.	.	.	.
Pechstein & Jenkins	1927	.	X	.	.	.	.	.	X	.	.	.	.	.	.
Kirkpatrick	1929	.	.	.	.	X	.	.	.	.	.	.	.	.	.
Goodenough & Anderson	1931	X	.	.	.	.	.	.	.	.	.	.	.	.	.
Brooks	1937	.	.	.	.	.	.	X	.	.	.	.	.	.	.
Boynton	1938	.	X	.	.	.	.	X	X	.	.	.	.	.	.
Strang	1938	.	X	.	.	.	.	.	.	.	X	.	.	.	.
Jensen	1938	.	X	.	.	X	.	.	X	.	.	X	.	.	.
Nagge	1942	X	X	X	X	.	.	X	X	X	.	.	.	.	.
Hurlock	1943	X	.	X	.	X	.	.	.	X	X	.	.	.	.
Jersild	1947	.	.	.	X	X	.	.	X	.	.	.	X	.	.
Breckenridge & Vincent	1949	.	.	.	.	X	X	.	.	.	.	.	X	.	.
Hurlock	1950	X	X	X	.	.	.	.	.	.	X	.	X	.	X
Millard	1951	.	.	.	.	.	.	X	.	.	.	.	.	.	.
Thompson	1952	.	X	.	.	.	.	X	.	X	.	.	.	X	.
Martin & Stendler	1953	.	.	.	.	.	X	.	.	X	.	.	X	X	X
Jersild	1954	.	.	X	.	.	.	.	.	.	.	.	X	X	.
Almy	1955	X	.	.	.	.	X	.	.	.	.	.	.	X	.
Hurlock	1956	.	.	X	.	.	.	.	.	.	.	.	.	X	X
Mussen & Conger	1956	X	.	.	.	.	X	.	X	.	.	.	X	.	.

Key to Table I:

- H = history and general remarks on the psychoanalytic movement
- Me = psychoanalytic method
- I = importance of early years
- F = theory of forgetting
- L = libido or sexual drive
- P = theory of personality (topographic and structural models), the unconscious
- Mo = theory of motivation
- C = conflict, maladjustment
- D = defense mechanisms, complexes
- Dr = theory of the dream
- Pl = theory of play
- Ch = childhood sexuality; psychosexual development
- In = infancy
- En = the environment, social values

TABLE 2
Percentage of Space by Decade Given to Various Topics
in Child Study and Child Development Textbooks

Decade \ Topic	Child study ^a	General description of stages of development (infancy, preschool, etc.)	Principles of learning	Biological functions ^b	Individual differences	Mental functions, language, intelligence ^c	Emotions	Moral development and discipline	Play, drawing, interests, imagination, dreams	Personality ^d	Mental hygiene, counseling, guidance, psychotherapy	Social development and socialization	Motivation	Other
1900-1909	20	17	-	52	3	7	-	-	-	-	-	-	-	1
1910-1919	20	11	2	24	1	17	4	3	11	2	-	-	-	5
1920-1929	29	3	8	30	5	10	5	4	2	1	1	-	2	-
1930-1939	11	39	2	11	2	8	4	1	3	9	3	3	1	3
1940-1949	8	12	4	16	2	17	9	2	7	8	1	6	2	6
1950-1957	9	22	2	11	1	10	6	4	5	7	2	20	-	1

^a Includes history, methods, general principles, applications to education, educational and vocational guidance.

^b Includes general biological principles, heredity, original nature, instincts, physical growth, health, physical disease.

^c Includes attention, sensation perception, memory, reasoning, curiosity.

^d Includes stages of personality, traits, adjustment, personality disturbances.

adopting a particular theoretical viewpoint of their own. Table 3 shows the number of index entries and the average number of pages devoted to the explicit discussion of Freud and other early important theorists, and the data suggest that it would be a mistake to expect to find a substantial discussion of any single individual's work. Only Gesell, who did not represent an independent theoretical position, clearly surpasses Freud in number of references. The increase through time of references to Freud corresponds to the data shown in Figure 1.

On the other hand, the inclusion of a wide range of material and the nontheoretical orientation of the authors are only a partial explanation of the slim coverage of psychoanalysis by the authors. If we restrict our attention to those chapters that were concerned with psychological matters as such, it becomes clear that many of the authors did not feel at home with the clinical traditions out of which psychoanalysis came. These sentiments are reflected in the remarks made by the authors about psychoanalysis, and it is our general impression that although several of the authors felt an obligation to mention psychoanalysis, presumably because of its emphasis on the early years, they did not wish to become very involved in an exposition of it and often proceeded with a decidedly ambivalent attitude. Three examples of this are the following.

Goodenough and Anderson (1931) say:

It is doubtful . . . whether the material regarding childhood experiences which has been elicited by the psychoanalytic method can be regarded as reliable. Both common observation and scientific experiment have shown that, with the passage of time, the memory of past events becomes extensively reorganized and modified as a result of later experience. . . . Moreover, even though their findings could be shown to be actually true for the individual cases studied, it would still be unsafe to assume that similar conditions would be found among the general population, since their data are derived almost wholly from the study of definitely abnormal cases. (pp. 15-16)

Strang (1938) writes:

Psychoanalysis has emphasized memories of past events as causes of present maladjustment, and undoubtedly has added a great deal to the sum total of knowledge regarding the hidden springs of conduct. The information a child discloses, however, is far from being spontaneous. Many of his associations are suggested directly or indirectly by the analyst and are interpreted from his viewpoint, which may not coincide with their real meaning to the child. (pp. 642-643)

Jersild (1954) says:

We give this brief mention to the Oedipus theory for two reasons. One is that the theory is so freely mentioned, both in scientific and popular literature, that we should pay our respects to it in this chapter. Secondly, and more importantly, we mention it because it contains, as do most of Freud's theories, a brilliant insight, even though it does not tell the full story. . . . But the theory of the Oedipus complex assumes a more rigid and uniform pattern of relationships between parents and children than the patterns that actually exist in one

TABLE 3
Tabulation of References to Prominent Theorists

Author	Date	Number of index entries ^a						Number of pages					
		Freud	Gesell	Hall	Pavlov	Thorndike	Watson	Freud	Gesell	Hall	Pavlov	Thorndike	Watson
Kirkpatrick	1907	-	-	4	-	3	-	-	-	1.3	-	.5	-
Tyler	1907	-	-	4	-	-	-	-	-	.8	-	-	-
Partridge	1910	-	-	-	-	-	-	-	-	-	-	-	-
Kirkpatrick	1911	1	-	1	-	-	-	.5	-	.5	-	-	-
Gesell & Gesell	1912	-	-	1	-	-	-	-	-	1.3	-	-	-
Tanner	1915	-	-	1	-	-	-	-	-	.5	-	-	-
Norsworthy & Whitley	1918	1	4	14	-	35	-	.1	.4	2.9	-	8.5	-
Waddell	1918	1	1	37	-	8	-	.1	.1	7.0	-	1.5	-
Averill	1921	-	-	5	-	3	-	-	-	2.1	-	.4	-
O'Shea	1921	-	-	3	-	1	3	-	-	.8	-	.1	1.4
Pechstein	1927	-	1	2	-	1	5	-	.6	.8	-	.6	6.5
& Jenkins													
Kirkpatrick	1929	2	2	2	-	2	-	.5	.5	1.2	-	.1	-
Strang	1930	-	19	3	-	12	9	-	2.6	.4	-	1.0	1.9
Goodenough & Anderson	1931	1	2	4	-	5	5	1.2	.4	1.6	-	.3	1.8
Brooks	1937	-	26	4	1	18	6	-	5.3	1.0	.2	3.5	1.8
Boynton	1938	-	10	4	-	4	1	-	3.3	.3	-	1.1	.4
Strang	1938	-	14	1	-	2	1	-	1.5	.4	-	1.2	.8
Jensen	1938	7	1	7	1	-	6	1.2	1.5	1.9	2.5	.5	.3
Reynolds	1939	-	2	-	-	-	1	-	.3	-	-	-	.5
Morgan	1942	-	13	5	-	4	5	-	1.1	.6	-	.3	1.4
Nagge	1942	14	37	11	10	13	20	2.7	7.8	2.5	2.5	1.8	2.1
Hurlock	1943	1	-	-	-	4	-	.1	-	-	-	.5	-
Pryor	1943	-	-	-	-	-	-	-	-	-	-	-	-
Gesell & Ilg	1946	-	9	-	-	-	-	-	-	-	-	-	-
Jersild	1947	12	31	-	-	-	4	.6	4.3	.1	-	.6	.5
Averill	1949	-	3	1	-	1	-	-	.2	.3	-	.3	.1
Breckenridge & Vincent	1949	2	39	2	-	5	2	.1	7.0	.2	-	1.0	.1
Hurlock	1950	8	42	9	-	2	20	1.7	10.7	2.0	-	.6	4.3
Millard	1951	2	14	-	1	8	6	.8	1.4	-	.5	.5	1.3
Thompson	1952	20	31	3	4	13	9	4.8	5.5	.5	.5	2.9	2.5
Almy	1955	2	8	1	-	-	1	-	3.2	.1	-	-	1.5
Hurlock	1956	8	61	5	-	-	-	.4	6.5	.4	-	-	1.4
Mussen & Conger	1956	16	22	3	-	-	11	.6	4.5	.5	-	-	1.2

^a Refers to separate index entries. Three of the 36 textbooks, namely, Reynolds (1951), Martin & Stendler (1953), and Jersild (1954), did not index authors.

culture as compared with others, in one family as compared with another family, and even within a single family in the relationship between a parent and one of his children as compared with his relationships with other children. There are many lines of observation and theoretical consideration which indicate that while the Oedipus complex might exist, as usually described, it cannot be taken for granted that it usually will exist. (p. 146)

On the other hand, other authors were decidedly more favorable to psychoanalysis, as in the following three examples. As might be expected, no simple uniform generalization about the attitude of the textbook authors toward psychoanalysis can be made.

Waddell (1918) says:

The psycho-analytic studies of Freud, Jung, and others have recently laid strong emphasis upon the moral bearing of the emotions of infancy and childhood. . . . An appreciation of the full significance of such subtle influences in the formation of character would no doubt often explain many a moral failure at adolescence or later. (pp. 212-213)

Pechstein and Jenkins (1927) say:

In sharp distinction to behaviorism, the new methods of psycho-analysis purpose to examine the innermost recesses of the mind, to uncover the complexes of the subconscious, and to furnish a proper outlet for repressed desires. This psycho-analytic technique has assisted in explaining as well as in improving cases of childish maladjustment. (p. 21)

Hurlock (1943) says:

Since Freud first called attention to the fact that many mental diseases of adult life are traceable to conditions that originated early in childhood, it has become increasingly apparent that good psychological foundations are as important as good physiological ones. . . . The importance of a good psychological foundation may be observed in the personalities of old people. The indelible impressions of childhood affect the whole life of the individual and color his outlook in old age as much as they did at the time the experiences occurred.

3.2 Educational Psychology

History of the movement. Although psychology as an experimental science is ordinarily said to have begun as an independent discipline in 1879 with the founding of Wundt's laboratory in Leipzig, informal psychological reflection on education has a much older history going back at least to Rousseau (1712-1778), Pestalozzi (1746-1827), and Herbart (1776-1841). Explicit attention to this older tradition was more or less part of the curriculum of the early normal schools in the country, the first of which was established in Lexington, Massachusetts, in 1839. By 1890 courses explicitly entitled *educational psychology* were offered by many teacher-training institutions; perhaps the highlight of this period was William James' delightful *Talks to Teachers* (1958), first given as lectures in 1892. A good sketch of the beginnings of educational psychology during the last two decades of the 19th century is to be found in Blair (1948).

A significant step toward the development of the educational psychology movement was taken in 1899 when E. L. Thorndike was given a faculty appointment at Teachers College, Columbia University. Under the leadership of Thorndike, Lewis Terman, and others, educational psychology had as its avowed aim the introduction of a scientific and experimental point of view in the study and solution of educational problems. This is already evident in Thorndike's 1903 text, in the preface of which he says: "This book attempts to apply to a number of educational problems the methods of exact science. I have therefore paid no attention to speculative opinions and very little attention to the conclusions of students who present data in so rough and incomplete a form that accurate quantitative treatment is impossible." In the years following, the relation between courses in educational psychology and courses in general or experimental psychology received intermittent attention in the literature. This attention properly reflected a tension between the scientific direction urged by Thorndike and others, on the one hand, and the opposite direction implicit in the courses in human growth and development which had a more clinical flavor (see, e.g., Hall-Quest, 1915; Woody, 1920; Bruce, 1949).

Textbooks. As in the case of the child-development movement, we examined the outstanding textbooks used by educational psychologists for evidence of explicit, implicit, and derived impact of psychoanalysis (already defined in Section 3.1).⁹ Figure 2 shows the trends over time. As in the case of Figure 1, the data show that influence increases with time, but to a lesser extent, and that the average number of pages remains quite small.

Psychoanalytic topics explicitly considered in the textbooks are shown in Table 4. Only 20 of the 44 texts examined showed such explicit influence. Freud's theory of personality (11 texts) received the most attention, but no topic is to be found in even a third of the texts examined.

Freudian topics of implicit interest do not appear in the table but it is undoubtedly the case that the theory of the defense mechanisms received the most of this kind of coverage. The first example of this occurs in Gates (1923), who devotes 13 pages to defense mechanisms, and substantial treatments are to be found in Jordan (1928), again in Gates (1930), Pressey (1933), Jordan (1933), Stroud (1935), Cuff (1936), and Commins (1937), as well as later works.

⁹ The 44 textbooks are Horne, 1906; O'Shea, 1906; Munsterberg, 1909; Pyle, 1911; Magnusson, 1913; Thorndike, 1913-1914, 1914; Starch, 1919; La Rue, 1920; Strong, 1920, 1922; Gates, 1923; Breitwieser, 1926; Starch, 1927; Tonks, 1927; Jordan, 1928; Pintner, 1929; Sandiford, 1929; Gates, 1930; Hollingworth, 1933; Pressey, 1933; Jordan, 1933; Stroud, 1935; Cuff, 1936; Commins, 1937; Sandiford, 1938; Judd, 1939; Mursell, 1939; Starch, Stanton, & Koerth, 1941; Jordan, 1942; Pressey & Robinson, 1944; Stroud, 1946; Peterson, 1948; Gates, Jersild, McConnell, & Challman, 1948; Simpson, 1949; Cole & Bruce, 1950; Guthrie & Powers, 1950; Ellis, 1951; Valentine, 1951; Mursell, 1952; Cronbach, 1954; Commins & Fagin, 1954; Garrison & Gray, 1955; and Jordan, 1956.

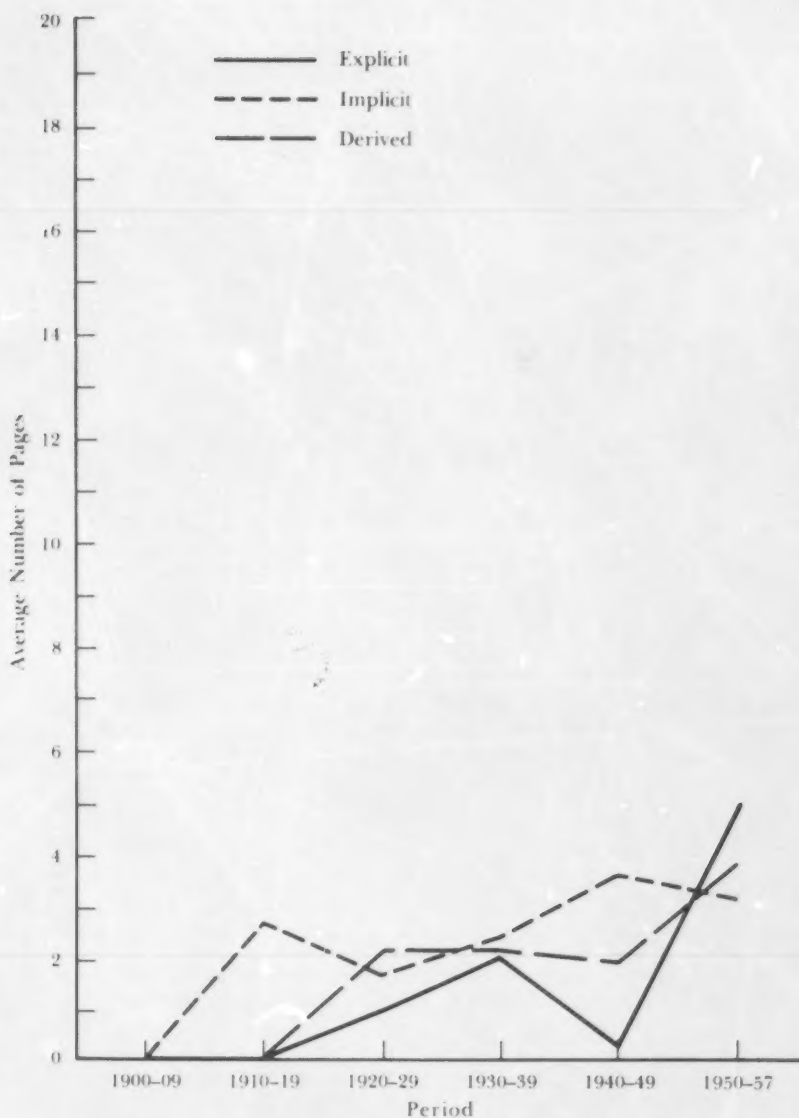


FIGURE 2. Evidence of psychoanalytic influence in educational psychology textbooks.

TABLE 4
 Psychoanalytic Topics Given Explicit Coverage
 in Educational Psychology Texts

Author	Year	Topics												
		H	Me	I	F	L	P	Mo	C	D	Dr	Ch	A	En
La Rue	1920										X			
Strong	1922				X									
Gates	1923					X	X		X	X				
Starch	1927							X						
Pintner	1929	X	X			X		X	X		X			
Sandiford	1929	X			X									
Gates	1930					X	X							
Stroud	1935		X				X			X				
Cuff	1936	X	X		X		X		X			X		
Commins	1937	X			X									
Sandiford	1938	X	X			X	X	X	X	X				
Mursell	1939		X		X	X								
Gates et al.	1948		X			X	X							
Cole & Bruce	1950					X	X	X	X	X		X	X	
Guthrie & Powers	1950			X		X	X	X						
Ellis	1951								X	X				X
Valentine	1951	X		X		X	X	X	X	X	X	X		
Cronbach	1954			X			X	X	X	X				
Commins & Fagin	1954				X			X						
Garrison & Gray	1955						X							
Jordan	1956			X				X						

Key to Table 4:

- H = history and general remarks on the psychoanalytic movement
 Me = psychoanalytic method
 I = importance of early years
 F = theory of forgetting
 L = libido or sexual drive
 P = theory of personality (topographic and structural models), the unconscious
 Mo = theory of motivation
 C = conflict, maladjustment
 D = defense mechanisms, complexes
 Dr = theory of the dream
 Ch = childhood sexuality; psychosexual development
 A = aggression
 En = the environment, social values

We were at first surprised at the extent to which the defense mechanisms were discussed in the texts, given the slight explicit treatment of psychoanalysis. We found, however, that the authors referred as frequently to mechanisms not discussed within the psychoanalytic literature as they did to the mechanisms discussed first by Freud and later by his daughter Anna Freud in her classic *The Ego and the Mechanisms of Defense* (1936/1946), as well as others; for example, alongside brief discussions of rationalization, projection, reaction formation, and so forth, we find such mechanisms as "logic-tight compartments," the "polyanna" syndrome, and the "sour grapes" syndrome. This suggests that consideration of the defense mechanisms was facilitated by their similarity to ideas coming from other sources.

Throughout, the texts were deficient in giving the proper attribution to Freud and psychoanalysis as the source of many of the defense mechanisms, a situation Hilgard (1949) has described well in his presidential address to the American Psychological Association:

The mechanisms of adjustment were the features of Freudian theory that we earliest domesticated within American academic psychology. They now have a respectable place in our textbooks, regardless of the theoretical biases of our textbook writers.

We took over the mechanisms when as a profession we were hostile to other aspects of psychoanalytic teaching. . . . We often gave only halting recognition to their psychoanalytic origins. Nearly all the mechanisms do in fact derive from Freud, Jung, Adler, and their followers. . . . Yet Gates' only mention of psychoanalysis was in some disparaging remarks about the "alleged adjustment by repression to the unconscious," an explanation of adjustment which he rejected as neither true nor useful. (pp. 374-375)

Representative sample remarks about Freud made by the textbook authors suggest that dislike of the sexual emphasis of psychoanalysis, as well as prevailing learning-theoretic conceptions of behavior, foreclosed the possibility of explicitly accepting psychoanalytic theory.

Strong (1922) says:

Freud claims that forgetting is also due to the fact that we want to forget, because the memory is unpleasant. . . . This type of forgetting may possibly be explained as due to interference. Take the case of the boy who is told that when he comes home he is to chop wood. There is here interference between spending the afternoon chopping wood and playing football. The latter is the stronger response due to habit and interest, and so "interferes" with the other response. Most failures to chop wood are sheer disobedience, but sometimes Freud is correct in saying that the duty is actually forgotten. (pp. 128-129)

Gates (1930) says:

Some students (*e.g.*, Freud and Watson) have concluded that the joyful emotions and the sex emotions are really the same. Usually such students assume that all forms of joyful emotions are really sexual in character. This notion seems to be incorrect. Everyday experiences and introspective study will disclose unique characteristics of forms of mirth, laughing, joyfulness and bodily exuberance that are unlike lust or any form of sex emotion. (p. 162)

And in 1936 Cuff says,

Freud has obviously introduced many bold hypotheses, some of which seem to be so valuable that they are now finding a place in our academic psychology. Several other concepts, however, have not been received with favor and are not likely to be accepted by leading American psychologists, who insist that it is unnecessary to introduce an instinctive drive which has presumably been suppressed. They also believe that the laws of habit formation can be used to formulate adequate explanations of maladjustments. Trow says the harm done by repressed desires is probably less than Freud supposes and that it may be more than is commonly supposed. Hence, it seems that psychoanalysis has made some contributions, but that it is not a panacea for all maladjustments. (p. 124)

Other comments show the concern with methodological adequacy that characterized the child-development texts. For example, Cronbach (1954) says, "Freud proposed that we distinguish two types of self-criticism—*ego* and *superego*. . . . Such a view is a bit too fanciful to study experimentally, but it helps us to distinguish moral learning from realistic learning" (p. 582).

Commins (1937) says, "The central point of the analytic doctrine, it seems, is whether the experience is unpleasant in retrospect, or more strictly, whether it would cause discomfort to the integrated personality. Experimental studies have not apparently always kept this in mind when attempting to put the Freudian doctrine to a test. . . . The clinical evidence offered by the psychoanalysts in support of their theory would seem to deserve recognition" (p. 456).

The commitment to scientific methodology seems to have gone a long way toward defining the subject matter of educational psychology since several studies, e.g., Hall-Quest (1915), Douglas (1925), Watson (1926), Worcester (1927), Weeks, Pickens, and Roudebush (1930), Emme (1942), and Blair (1948), show that from the first there was little agreement about how the field was to be defined from a substantive standpoint. The authors comment on the wide diversity of topics included in the texts, and we reprint one such analysis as Table 5. (The percentages shown in Table 5 do not sum to 100 for each text because some topics were treated by Blair as overlapping, and the full content of the text was not always accounted for.)

However, in surveying 83 English-language educational psychology textbooks by decades, Gates, Jersild, McKillop, Rivlin, Shoben, and Watson (1956) report a steady and significant increase in the percentage of texts covering the topics of mental hygiene, unconscious motivation, and counseling, guidance, and psychotherapy, as well as a uniform increase in the number of pages devoted to these topics. They point out that attention to these matters is the outstanding development during the four decades they review. Reproducing their table as Table 6, we see that the data are uniformly increasing in time, with one minor exception. As their own extensive references make apparent, it is reasonable to attribute some

TABLE 5
Percentage of Space Given to Various Topics in Each of 13 Textbooks in Educational Psychology (Blair, 1948)

Topic	Bruce & Freeman	Cruze	Gates	Hartmann	Jordan	Pressey & Robinson	Skinner (Elementary)	Skinner	Sorenson	Starch	Stroud	Witherington	Woodruff
The Field of Educational Psychology	0	6	1	6	3	2	7	3	0	1	4	3	2
Heredity and Environment	7	7	1	2	8	9	1	1	12	1	9	2	0
General Growth and Development	6	2	3	7	1	22	4	4	5	2	7	5	5
Child Psychology	23	4	3	5	1	6	2	2	1	5	1	1	1
Adolescent Psychology	6	3	3	1	1	5	0	1	1	6	0	0	0
Personality	1	6	13	6	5	0	4	3	0	0	8	5	3
Mental Hygiene	6	5	4	1	1	0	5	4	4	2	0	4	3
Intelligence and Its Measurement	1	5	3	9	8	2	10	7	5	1	10	9	15
Individual Differences	2	2	7	0	6	7	5	6	6	2	7	2	7
General Nature of Learning	26	0	5	1	2	5	4	9	8	0	6	2	24
Factors Influencing Learning	5	9	8	11	8	9	4	1	31	5	8	4	11
Transfer of Training	0	11	4	3	10	12	4	4	5	1	6	5	3
Reasoning and Problem Solving	3	1	4	5	1	0	4	6	0	1	1	5	4
Tests and Measurements (Learning)	6	3	9	1	4	5	8	4	9	5	5	6	10
Tests and Measurements (Personality)	0	3	1	1	4	0	0	0	0	1	0	1	1
Statistics	0	0	1	0	4	0	0	0	1	1	0	0	0
Schools of Psychology	0	1	1	0	0	0	1	4	0	0	1	1	0
Psychology of School Subjects	0	0	1	9	4	4	6	5	0	46	10	17	0
Basic Theory of Motivation	1	4	2	0	0	0	1	1	5	1	2	1	2
Motivation (Devices in Teaching)	0	1	1	5	3	3	0	2	5	6	6	2	2
Psychology of Exceptional Children	0	4	2	1	3	4	5	6	0	1	0	0	3
Educational and Vocational Guidance	3	5	4	1	0	0	8	3	0	6	0	0	2
Mental Health and Personality of the Teacher	2	1	4	4	0	1	4	4	1	1	0	1	4
Social and Emotional Development	14	4	15	12	2	6	8	13	13	1	3	5	14
Development of Originality and Creativity	1	0	0	5	0	0	4	3	0	1	0	0	0
Adult Education	0	0	0	2	0	0	0	0	0	0	0	5	0

TABLE 6

Trends in Amount of Space Devoted by Textbooks in Educational Psychology to the Dynamics of Personality (Source: Gates et al., 1956)

Period	Proportion of texts discussing the topic			Average number of pages devoted to the topic		
	Mental hygiene	Unconscious motivation	Counseling, guidance, psychotherapy	Mental hygiene	Unconscious motivation	Counseling, guidance, psychotherapy
1920-29	46%	30%	23%	7	4	1
1930-39	75	63	46	17	6	8
1940-49	100	86	79	39	21	23
1950-56	100	95	95	53	46	19

of this development to psychoanalysis, but their definition of unconscious motivation is much broader than the psychoanalytic framework we have used in determining the data relevant to Figure 2.

Student and teacher attitudes. It is probable that much of changed emphases in the texts can be attributed to student and teacher pressure. As early as 1911, O'Shea reported in the second volume of the *Journal of Educational Psychology* that a survey showed graduate students in education felt that "physics was as close to the problems of training children as this work in laboratory psychology" (p. 331). A similar and highly influential article was published later by Watson (1926), who sampled some faculty members of Teachers College, a group of experienced teachers and administrators, and a group of undergraduates who had not taught but were preparing for teaching, concerning what should be taught in educational psychology. Watson concludes that "the apparent concurrence of opinion would place major emphasis upon problems of emotion and personality adjustment, problems of original nature and heredity, and problems of general teaching method. Little emphasis would be placed upon the psychology of the acquisition of skill, or the psychological problems involved in home-making and pre-school child training, or the various schools and theories in psychology" (p. 596). Two years later in a survey of the opinion at a teacher-training institution concerning the relative value of subject matter in educational psychology to the elementary-school teacher, Hertzberg (1928) found that, in agreement with Watson's survey, problems concerning personality adjustment were felt to be a subject of major importance. When Cuff (1935) conducted a survey of the attitudes of educational psychology teachers (97 in number)

concerning the appropriateness of topics to be included in the course in educational psychology as reflected in chapter headings of texts, he found that the teachers seemed to agree with each other more closely than the textbooks did with each other. Also, more teachers felt that the contents suggested by such a chapter heading as "Incentives and Motives" should be included than one whose subject matter was indicated by the title "The Measurement of Learning."

3.3 *Mental Hygiene*

History of the movement. There is a symbolic if not real date for the beginning of the mental hygiene movement in this country. On May 6, 1908, a small group of people gathered in New Haven, Connecticut, at the invitation of Clifford Whittingham Beers, to organize the Connecticut Society for Mental Hygiene. Beers had already written *A Mind That Found Itself*, a remarkable autobiographical account of his recovery from insanity, and much of the early work of this Committee and its staff was concerned with the conditions of the mentally ill in institutions and the mentally defective.

The first official interest in mental health education came in 1917, when the quarterly *Mental Hygiene*, the publication of the National Committee for Mental Hygiene, first appeared. This journal, as well as the various pamphlet publications of the National Committee for Mental Hygiene, served as the introduction to the subject for educators and other community leaders. By the time of William A. White's (1930) presidential address to the First International Congress on Mental Hygiene in Washington, D.C., on May 6, 1930, the positive aspects of mental hygiene and its close relation to the concerns of preventive medicine were being emphasized. It was in this context that White stressed the educational implications of mental hygiene and its usefulness to the educational programs in the schools. The impact of the mental hygiene movement on education, which had become a serious systematic influence by the late 1920s at least, is described well in the following quotation from Ridenour (1961).

When E. K. Wickman's book *Children's Behavior and Teachers' Attitudes* appeared in 1928, his charts were a revelation. Table after table showed that what teachers regarded as signs of maladjustment in children, psychiatrists passed off as normal behavior, and vice versa. At the extremes ("most serious" and "least serious"), there was almost no overlap. Teachers were troubled by aggressiveness in all its forms and by any behavior which contravened the discipline of the classroom, now often referred to as "acting-out behavior." Psychiatrists were more concerned about signs of withdrawing and evidence of a child's inability to get along with his peers. All this was "news," both to the teaching profession and to the recently created clinic teams (Mr. Wickman himself was at that time a psychologist with the Commonwealth Fund demonstration clinics) who were just then beginning to concern themselves about the adjustment of children in school. Wickman's book was one of the milestones in the

long educational process of teaching teachers about mental health in the classroom. (pp. 88-89)

A great deal has been written about the mental hygiene viewpoint relevant to classroom teachers' attitudes. Bernard (1952, Chap. 1) gives a quotation from Ryan (1945, p. 378) that provides an especially succinct summary of the viewpoint.

Mental hygiene has to do primarily with the development of more wholesome human relationships. It means applying to everyday living what has been learned with regard to the behavior of human beings. On the negative side, it has been concerned with more humane and intelligent care and treatment of the mentally ill. Positively, it has meant the early discovery of mental and emotional difficulties, the prevention of as much threatened serious illness and disorder as possible, and the provision of healthful living for the greatest conceivable number of individuals as members of a modern society. For teachers, social workers, and others whose daily activities have to do with children, youth, and other human beings, mental hygiene involves not only a better understanding of human growth and development with respect to those with whom they work, but also an understanding of the worker's own personality and his own relationship to his family, associates, school, and community. (Bernard, 1952, p. 15)

Courses. The growth of the movement in the schools has been charted in several articles providing quantitative data on both the number and the kinds of courses taught. The earliest study we have found is Burnham (1924)—considering the 1912 date of the founding of the National Committee for Mental Hygiene, one would not expect to find an earlier survey.

Burnham reports that during the academic year 1919-20, a questionnaire under the sponsorship of the National Committee for Mental Hygiene was sent to "the public and private normal and training schools throughout the country" (p. 19). Replies were received from 175 normal and training schools with an enrollment of approximately 40,000 students, about 80% of whom were women. There were nine main questions, with some of the questions consisting of several subordinate questions. The first six are most relevant to the concerns of this chapter and so we quote them here (Burnham, 1924, p. 19), along with summaries of the responses.

1. Is any regular course in mental hygiene given in your school? If so, what is the character of the course and how much time is devoted to it?

Approximately 78% of the schools reporting state definitely that no regular courses in mental hygiene were given.

2. Is any *special* course in mental hygiene given in your school? If so, what is the character of the course and how long a time is devoted to it?

Approximately 73% of the schools reporting indicate that no special courses in mental hygiene were given.

3. Is any part of the field of mental hygiene taught in connection with the courses in psychology? If so, what topics are presented and how much time is devoted to the subject?

A majority of the schools, approximately 59%, report that some aspects of mental hygiene were taught in connection with courses in psychology.

4. Is any part of mental hygiene taught in connection with the course in child study in your school? If so, what topics are taken up and how much time is devoted to the subject?

Approximately 45% of the schools report that work in mental hygiene was given in connection with a course in child study. Burnham does not indicate whether the data show that every school reporting also had a course in child study.

5. Is any part of mental hygiene taught in connection with school hygiene in your school?

A majority of the schools report that no part of mental hygiene was taught in connection with school hygiene, but approximately 35% indicate that some part was taught in connection with the school hygiene course.

6. Is any part of mental hygiene taught in connection with any other courses in your school?

Again, a majority report that no part of mental hygiene was taught in connection with other courses but approximately 23% indicate some instruction in mental hygiene in connection with other subjects, such as social science and biology.

Burnham also asked various schools to include comments, and the special remarks made during these early years about the courses indicate little influence of psychoanalysis. Rather, there is a good deal more attention given to mental tests, problems of the mentally retarded, and the grosser symptoms of psychiatric disturbance. The response from the Department of Physiology, State Normal School, Ypsilanti, Michigan, did, however, indicate in the outline sent that material on Freud and neuroses was taught.

Burnham also included a bibliography of representative books used in various normal schools, and of his list of 27 texts, probably only the classic book by William A. White, *Principles of Mental Hygiene*, had much to say about psychoanalysis. It is not possible to judge exactly how representative Burnham's sample of 175 schools was, but given the large number, it may be taken as reasonably so and it is doubtful if a larger or more representative sample would have substantially changed any of the conclusions.

Ten years later a similar survey was conducted by Benson and Alteneder (1931). Their questionnaire was sent out in January 1930 to the 295 teacher-training institutions listed in the Educational Directory for 1929. These institutions included 137 teachers colleges, 71 state normal schools, 27 city public normal schools, and 60 private teacher-training

institutions. Of the 295 institutions, 239 (81%) returned information. In the 11-point questionnaire, the questions of most interest here were the following (p. 226):

1. Does your curriculum include a regular course or courses in mental hygiene?

Twenty-two percent of the institutions responding report a regular course in mental hygiene.

2. Is the subject of mental hygiene presented in connection with other courses, such as educational psychology, child psychology, general hygiene, etc.?

Over 80% of the institutions reporting indicate that the subject of mental hygiene was presented in connection with other courses, mainly courses in psychology.

3. Have you special lectures in mental hygiene during the year?

Approximately 24% of the institutions reporting indicate that they had special lectures in mental hygiene. These lectures covered various concerns of the field, such as "mental health, mental disease, emotional maladjustments, sex problems, and so forth" (p. 228).

6. Have you a psychiatrist in your institution, for the help of students, or for the help of children in a practice school? If not, is there a member of your faculty capable of giving mental-hygiene help?

Less than 6% of the institutions reporting indicate that a psychiatrist was connected with an institution or was available for consultation and help. On the other hand, over half of the institutions indicated that a member of the faculty was capable of giving mental hygiene help.

7. Have you a mental-hygiene clinic in your institution, or accessible to your students, and if so, what is the equipment?

Less than 14% of the institutions reporting indicate that a mental hygiene clinic was accessible to the students. Some of the clinics that were mentioned were part of the teacher-training institutions; others were child-guidance clinics.

Question 10 dealt with the textbooks in mental hygiene used by the institutions. The three books most widely cited were Burnham (1924), cited by 44 institutions, La Rue (1927), and Morgan (1924), both cited by 26 institutions.

Benson and Altene'ier (1931) present an interesting graph showing the distribution by states of teacher-training institutions that provide regular courses or special lectures in mental hygiene. Only two states, Florida and Wyoming, as well as the Commonwealth of Puerto Rico, show no such institutions. The four states with the most extensive activities are Illinois, Massachusetts, New York, and Pennsylvania. The authors make a detailed comparison of the results of their survey with that of Burnham's ten years earlier. Although we shall not summarize that comparison, it is

evident that the educational role of courses and lectures in mental hygiene had increased considerably from 1920 to 1930.

Ten years later a third but more superficial survey was made by Bernard (1940), who examined the course catalogues of 178 colleges. The catalogues were not selected randomly but apparently were collected until the desired number of 75 colleges offering courses in mental hygiene had been identified. Bernard then compared the findings for the same colleges ten years earlier by examining their course catalogues for 1929-30. Of his sample of 75 colleges offering courses in mental hygiene in 1939-40, he found only 28 schools, or approximately 37%, that had offered such courses ten years earlier. He notes that his sample of 178 colleges constitutes about 27% of the total number of American colleges and universities listed in the 1938 *Educational Directory of the United States Office of Education*. Thus his results, though obtained by rather unusual sampling methods, give a reasonably representative picture.

Attitudes of teachers. More indicative of the impact of psychoanalysis on the mental hygiene movement are the various studies of teacher attitudes made during this period. Using more than 500 teachers as respondents, Wickman (1928) conducted a classic investigation that revealed marked differences between the interpretation of children's behavior problems by teachers and by mental hygienists. Although teachers rated as most serious transgressions against authority, dishonesty, and sexual immoralities, the mental hygienists rated these items as relatively minor in importance but, as more serious, unsocial, withdrawing, or regressive behavior. The hygienists also rated as fairly serious cruelty, temper tantrums, and truancy, each of which was not so rated by the teachers. The methodological excellence of Wickman's study set a high standard for a number of related early studies (Yourman, 1932; Peck, 1935; Ellis & Miller, 1936), and the mere fact that these researchers were permitted to ask for responses from teachers in the school systems is indicative of the impact of the mental hygiene movement in the American school setting. Studies of this kind would have simply been unthinkable 50 years earlier.

In 1942, Mitchell conducted a study similar to Wickman's primarily to determine whether teachers and mental hygienists had come more into agreement in their estimates of the relative seriousness of student behavior problems. Broadly speaking, his results showed that though the mental hygienists had become somewhat more conservative, the teachers had changed more than the mental hygienists in their basic attitudes. The results of Mitchell's study were confirmed by Stouffer (1952), who found the most marked differences between the two groups (teachers and clinicians) to be in the ratings given to disobedience, sexual activity, profanity, smoking, obscenity, and unsocial behavior. Unhappiness, depression, unsociability, and withdrawal had moved closer to the top of the list in

seriousness, according to the teacher ratings.

Sparks (1952) tested out the independent variables of experience and education for relative effect on teachers' attitudes. He found that teachers with more education were closer to clinicians in their appraisal of children's behavior but that amount of experience had little effect. Schrupp and Gjerde (1953), using data from more than a hundred San Diego teachers, attempted a detailed replication of Wickman's study. They found that the attitudes of 1951 teachers, the year when they obtained their data, were much more in agreement with clinicians than was true for the 1927 teachers; they also point out that teacher attitudes can be expected to differ from ideal criterion attitudes of mental hygienists because the teachers are inevitably concerned with temporary but disturbing behavior in the classroom. They conclude their article with the remark that in view of the continuing disagreements between the attitudes of teachers and clinicians, there remains a need for the development of what they call a "mental hygiene viewpoint" in teacher education, both at the pre-service and inservice level.

Hunter (1957) undertook a detailed study to compare classroom teachers in 1955 with those studied by Wickman in 1926-1927. One of the more striking changes he found was in the ranking of sexual activity. Teachers in the Wickman study ranked heterosexual activity and masturbation first and third in seriousness; the teachers in Hunter's study moved these ranks to 10th and 28th, respectively. More generally, Wickman reported a correlation coefficient of $-.11$ between the rank order of traits in terms of seriousness by the mental hygienists and by the teachers. Using the same mental hygienists' ranking that Wickman did, he was able to report a correlation coefficient of $.22$ between the teachers' responses and the hygienists'. Without entering into the details of the sample size, geographical location, and other variables, it still seems reasonable to conclude that a significant change in attitude had occurred between 1926-1927 and 1955. We would be inclined to attribute some of the marked changes in attitude toward sexual behavior to the indirect influence of psychoanalytic ideas as they increasingly permeated the culture during the period covered by this review.

These findings are broadly consistent with several studies that attempted to assess how much teachers know about mental hygiene. An excellent example is one by Hobson (1937), who conducted an extensive survey of teachers in North Dakota organized along the following lines. First, a list of mental-hygiene principles was constructed. The author began with a list of 160 principles and sent them to 12 individuals working in mental hygiene, so as to obtain a reliable subset of 42. A simple imaginary case exemplifying each of the 42 principles was constructed, and judgments about some aspect of the case were given as test items to teachers. The answers to the test items were validated by a panel of 16 judges; any item

on which less than 87% of the judges agreed was discarded, that is, 14, 15, or 16 judges were required to agree on the correct answer.

Among the 42 principles, we would identify five or six as having possible antecedents in psychoanalytic theory. We emphasize that the language of psychoanalysis was not used explicitly and it would also be reasonable to claim that other theoretical sources were responsible for the inclusion of these principles. All the same, we think there is some *prima facie* case for considering them evidence of the influence of psychoanalytic theory on the mental hygiene movement. We quote the six in which we think this influence is probably the most salient (Hobson, 1937, pp. 233-236).

2. The conditioning forces of early years are the most significant forces in determining emotional patterns of behavior, which patterns tend to persist in adulthood, being the principal determinants of maladjustment or adjustment—i.e., personality.
6. Many neurotic conditions result from a feeling of insecurity.
8. Sometimes the fearful child tends to compensate by developing an egocentric, aggressive, overbearing attitude.
9. Fear of ridicule and the hostility of others may lead to all sorts of reactions, as rationalizations, excuses, lying, running away, etc.
12. The adult's reaction to the child's natural curiosity about sex may result in having the child carry ever after a sense of guilt that attaches to anything that has to do with sex.
40. In child development, success is more likely to follow treatment of the underlying pathology of the behavior disorder than treatment of the symptoms of the behavior maladjustment.

In informal and summary fashion Symonds (1941) discusses his views, based on his own teaching experience, on the value of courses in mental hygiene for the personality adjustment of prospective teachers. His discussion of projection and repression in the consideration of case examples is psychoanalytic in spirit.

Weinreb (1953) reports an effort to provide special lectures for teachers in which the genetic development of personality from the psychoanalytic point of view was presented. This was done over a number of years in Massachusetts in towns like Leominster, Gardner, and Needham. Considering the fact that the teachers who were the students in the course were almost uniformly naive about psychiatric or psychoanalytic concepts, the author's remarks about the development of the teachers' attitudes are of some interest. "The course usually started out with pronounced reserve, followed by a rise in hostility to the author, the peak of hostility being reached around the eighth session, when the Oedipal conflict was discussed. . . . After this session there always followed a session in which various members of the group would come out with evidence from their own experience confirming what had been brought

out regarding the Oedipal conflict, as if they had discovered the answer to an age-old riddle" (pp. 291-292).

Falick, Peters, Levitt, and Rubenstein (1954) describe in informal terms an extensive effort in a small town in Michigan to raise the understanding and sophistication of teachers about the psychological problems of their students. The authors have to a large extent a psychoanalytic orientation and have a number of interesting things to say about the resistances they encountered in teachers as they became aware either of their students' problems or of their own problems that were interfering with their ability to accurately perceive their students. For example, many of the teachers were able to enter into detailed discussion of aggressive drives in children but were not able to handle frank and detailed discussions of sexual drives, especially in those students whose behavior manifested some homosexual component.

Textbooks. We examined 22 texts used in courses on mental hygiene. Figure 3 shows the results of our analysis arranged chronologically by decades, starting with 1900. The trends are the same as those for child study and child development (Figure 1) and educational psychology (Figure 2), but the much greater average number of pages showing the explicit influence of psychoanalysis on mental hygiene is clear evidence of greater impact. In comparison to Figures 1 and 2, two other differences are noticeable. One is the much earlier rapid increase in explicit influence in the period 1910-1919, and the other is that explicit influence exceeds implicit or derived influence in all but the first period.

A systematic picture of the emphasis given to various topics in the 22 textbooks we examined is shown in Table 7. As would be expected, the general content of the texts on mental hygiene is considerably closer to psychoanalysis than those on child development or educational psychology, but as is also evident, some topics judged important on the basis of their coverage are quite separate from psychoanalytic concerns. Moreover, topics like personality and social development are presented from several different theoretical viewpoints, usually with a deliberate attempt to be eclectic in approach.

Of the textbooks we examined, only three ignored psychoanalysis, namely, Averill (1928), Cutts (1941), and Bernard (1940). Of the other 19 texts examined, there are many with more pages mentioning psychoanalysis than is typical of the texts on child study, child development, or educational psychology, and six with complete chapters on the theory. Put another way, there is probably no text in the other disciplines that shows the extensive presentation of psychoanalytic ideas characteristic of the texts on mental hygiene.

Five fairly late examples in the period we are covering provide ample evidence of this. Wallin (1949) states explicitly in the preface to the second edition of his book (the first edition was dated 1935) that "the material on

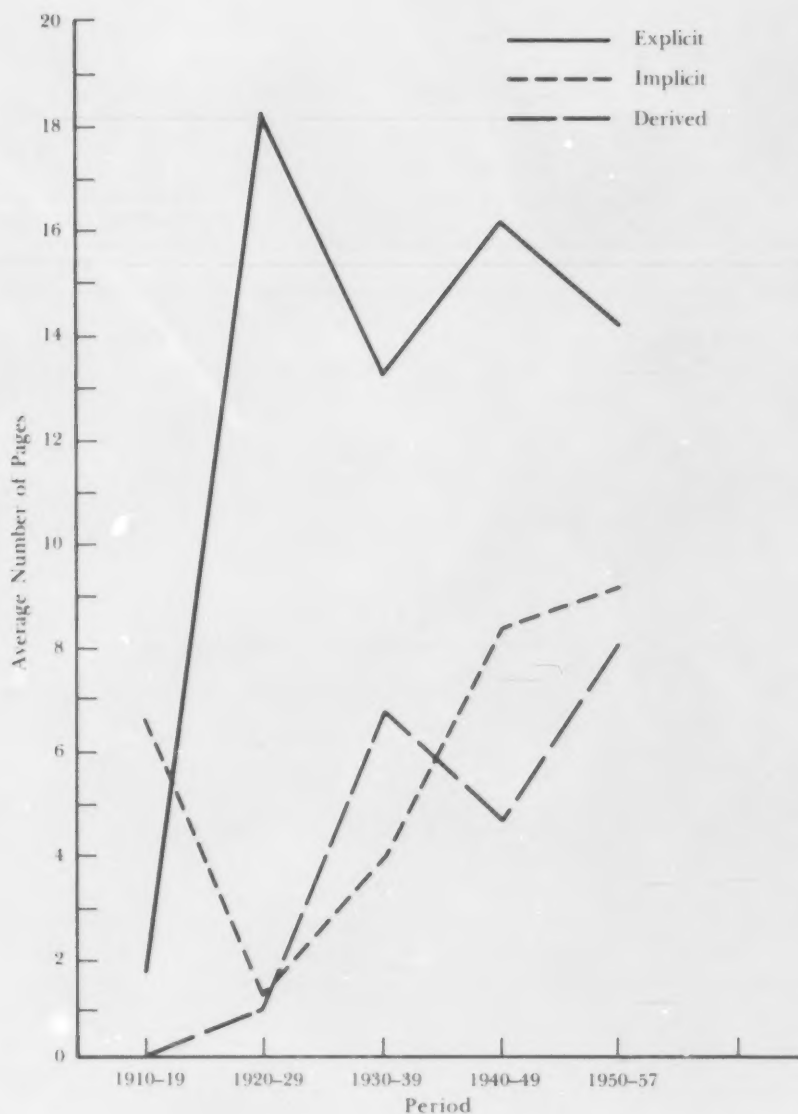


FIGURE 3. Evidence of psychoanalytic influence in mental hygiene textbooks.

TABLE 7
Percentage of Space Given to Various Topics in Mental Hygiene Textbooks

Author	Year	Fundamental needs and motivation	Personality and social development	Personality adjustment and maladjustment, including diagnosis	Mental hygiene and the family	Mental hygiene and the school and community	The teacher	Therapy and guidance, including case studies	Individual differences	Learning	Organic bases of personality and physical hygiene	Mental hygiene movement, including methods and general principles	Other
Wallin	1914	-	-	-	-	21	-	7	26	-	20	26	-
White	1919	17	39	-	17	-	-	-	-	-	-	14	13
Burnham	1924	-	22	11	-	10	-	26	5	21	-	5	-
Averill	1928	-	-	14	7	34	-	7	15	7	5	11	-
Groves & Blanchard	1930	-	19	14	-	15	-	-	-	-	-	21	31
Symonds	1934	18	14	-	-	8	13	31	-	-	-	16	-
Sherman	1934	-	46	46	-	2	-	-	-	-	-	-	6
Shaffer	1936	12	21	31	-	-	-	-	5	7	23	1	-
Ryan	1938	-	-	-	-	54	24	11	-	-	-	11	-
Averill	1939	-	-	-	-	-	100	-	-	-	-	-	-
Tiegs & Katz	1941	-	10	55	4	5	-	8	-	-	7	10	1
Cutts & Mosely	1941	-	24	3	7	34	20	11	-	-	-	1	-
Fenton	1943	-	7	17	5	8	21	22	-	-	-	18	2
Carroll	1947	34	13	16	-	9	-	14	8	-	-	6	-
Wallin	1949	-	-	45	-	4	-	14	9	-	-	28	-
Thorpe	1950	5	18	38	6	18	-	5	-	-	5	4	1
Redl & Wattenberg	1951	-	18	19	6	11	18	18	-	-	-	8	2
Carroll	1951	38	-	29	-	8	-	5	14	-	-	6	-
Bernard	1952	16	10	10	-	15	21	20	-	-	-	11	2
Lindgren	1954	13	32	11	-	14	19	6	-	-	-	5	-
Carroll	1956	33	-	20	-	13	-	7	15	7	-	5	-
Shaffer & Shoben	1956	-	35	39	-	-	-	6	-	7	-	12	1

psychoanalysis occupies the last five chapters. My experience has been that students are keenly interested in this subject, and that is why it has been retained in the new edition" (p. viii). Indeed these last five chapters occupy over 100 pages of the text. The author's attitude is well illustrated by the section heading "Importance of Acquaintance with Psychoanalytic Concepts in Spite of Their Controversial Character" (p. 474).

The second book that gives a systematic exposition of psychoanalytic ideas is the one by Shaffer (1936). Shaffer prides himself on being a strict advocate of "objective" psychology and, as would be expected, is hostile to psychoanalytic thinking. He says in his preface: "Especial care has been taken to avoid the use of psychoanalytical conceptions, either explicitly or implicitly. The hypotheses of 'the unconscious' and of 'complexes' may have been of some temporary service in the past, but the author feels that they are no longer useful. Objective psychology is now prepared to attack the problems of human adjustment without the assistance of these doubtful assumptions. Because of their historical importance, the principal theories of psychoanalysis are presented rather fully in a separate chapter" (p. ix). In Chapter 14 the author does give a systematic exposition, and the account is rather detailed and relatively accurate. He discusses in some detail the concepts of id, ego, and superego, noting that he has attempted to draw from the later writings of Freud as well as the earlier works. He discusses the oral, anal, and genital stages of development, and the Oedipus and castration complexes.

As in the case of the educational psychology texts, it is intriguing to find that in other chapters Shaffer discusses extensively the defense mechanisms in a positive and detailed way. It is clear that in his thinking the classical defense mechanisms do not come from psychoanalysis but from other sources in the psychological literature, although he does not suggest what these alternative historical sources are.

Somewhat surprisingly, the second edition of Shaffer's book, Shaffer and Shoben (1956), adopts a much more favorable attitude toward psychoanalysis, and the strident tone with which he emphasized the great merits of objective psychology is dropped. In the new preface, the authors say, "The account presented here is derived from three great movements in twentieth century psychology—the experimental study of behavior, the psychoanalytic approach to psychopathology, and the conceptions of interpersonal relationships developed by social psychology, sociology, and anthropology" (p. ix). Later, the authors say, "The experimental sources have been leavened by the authors' own clinical experiences, and by the impacts of psychoanalysis. The great contributions of Sigmund Freud and of the movement he fathered have been slow to merge with the stream of general psychology, perhaps because differences of language long obscured essential similarities. The rapprochement between psychology

and psychoanalysis has been due in large part to a number of contemporary psychologists whose theories and experiments have brought the clinic and the laboratory into a close and fruitfully reciprocal relationship" (p. xi).

Even the discussion of defense mechanisms shows an impact of psychoanalytic thinking in a more explicit way. The terminology adopted is practically taken from standard psychoanalytic sources: the mechanisms of compensation, identification, reaction formation, rationalization, and projection are mentioned in the order stated here. The literature references show a similar movement toward a psychoanalytic framework, and the chapter explicitly concerned with psychoanalysis has been revised in tone in the same way. The authors now say in the opening paragraphs of the chapter, "Psychoanalysis was an early approach to the problems of troubled people, and a knowledge of it gives historical perspective. Also, psychoanalytic concepts have had a great impact on our entire culture, on literature and art as well as on psychology, so that every liberally educated person needs to know something of them" (p. 448).

In the first edition of Shaffer's book, the final section of the chapter on psychoanalysis, providing an appraisal or evaluation, was extremely critical and negative in tone. The second edition adopts quite a different viewpoint. For example, it is stated early in the final section, "Sigmund Freud was without doubt one of the world's authentic geniuses. Few modern men—Newton and Darwin are other examples—have had so broad an influence on the thinking of our culture" (p. 472). The authors repeat some of the criticisms of psychoanalysis as a science but also now emphasize the developing rapprochement mentioned in the preface. As they put it, "To other men falls the task of sorting Freud's great hunches into the verified and the unverifiable" (p. 478). We see in this second edition of Shaffer's book the kind of influence that psychoanalysis has accumulated by the end of the period dealt with in this chapter. Its pervasive and extensive impact, not only on psychology but on the whole of our culture, is recognized by a psychologist who prides himself on the use of experimental methods and his affiliation with 20th-century theories of learning.

The third of the five texts mentioned above is Rivlin (1936). There is a chapter on psychoanalysis, written in a generally favorable tone. The discussion is not as detailed as in Shaffer, but many psychoanalytic concepts are discussed at least briefly. There are also sections on Adler and Jung. As in other books on mental hygiene, the discussion of defense mechanisms is placed in other parts of the volume and they are not identified as a psychoanalytic contribution.

The fourth book containing at least one chapter on psychoanalysis is Thorpe (1950). The author's attitude toward psychoanalysis and toward

other matters is expressed in the following passage in the preface: "The point of view . . . is an eclectic and flexible one. Although an effort has been made to present a consistent picture of the causes, conditions, and correction of mental ill-health, the author has endeavored not to take sides on debatable issues. Because of widespread interest in its proposals, a chapter on psychoanalysis has been placed near the beginning of the book, thus providing a basis for later comparison of its postulates and claims with those of other schools of thought in the mental health field" (p. vi).

The fifth book is Redl and Wattenberg (1951), which does not contain a separate chapter on psychoanalytic theory but has an extensive appendix of psychoanalytic terminology. Broadly speaking, the book is typical of many in this area in the sense of being eclectic and not deeply oriented toward any particular theoretical standpoint. The same book has a 1959 second edition (which is slightly outside the period covered by our review) and it is worth noting that in the preface to the second edition the authors note that three new chapters have been added, the first of which "deals with unconscious motivation, conflict, and the extent to which children have conscious control over their behavior" (p. xiii).

Despite the increasing acceptance of Freudian theory, it remains the case that the discussions of the defense mechanisms are usually separated almost completely from those of psychoanalysis as such. Since many of these same texts refer frequently to a group of early texts and treatises written in the earlier part of the century (roughly in the period 1911-1918), we could not forbear investigating the analysis of the defense mechanisms in these earlier books and the extent to which the authors attributed the source of the concepts to Freud. We restrict our comments here to the books by one of the most prominent people in this period, William A. White, who was probably the most important medical figure in the early history of mental hygiene.

White published three widely read books (1911, 1916, 1917). All three of them explicitly acknowledge Freud as the source of the theory of defense mechanisms. The 1916 book, in fact, which is entitled *Mechanisms of Character Formation*, has as its subtitle *An Introduction to Psychoanalysis*. The widely cited 1917 book, *The Principles of Mental Hygiene*, has as its third chapter "Mental Mechanisms," and it is an elementary-textbook discussion of the concepts more thoroughly developed in the 1916 book; in this chapter he mentions specifically projection, identification, and rationalization.

White's attitude and sense of indebtedness to psychoanalysis are vividly conveyed in the opening paragraph of the preface to his 1916 book. We cite part of the paragraph here.

Certainly the field of mental medicine has benefited practically not at all as a result of all the years of laboratory psychology. A man seems to have been considered by the psychologist as an object of experiment and rarely as a human being in a social environment. True, the behaviorists may change all this but in the meantime a new psychology has come into existence, borne of the sufferings and the heart aches of the mentally ill—the psychology which is called psychoanalysis and, no matter what the remote history of events preceding its birth, properly also bears the name of its real creator, Prof. Sigmund Freud of Vienna—Freudian. This is a psychology which had its origin in trying to help sick people, in trying to alleviate their sufferings and from the very first dealt with men and women in the raw, as they really were. (p. v)

3.4 *Progressive Education*

What is called the progressive education movement began in the latter part of the 19th century, although no definite date of origin can be set for a diffuse and complex activity of this kind; a conventional date that has symbolic value is the founding by John Dewey of the Laboratory School at the University of Chicago in 1896. An excellent critical history of the progressive movement has been given by Cremin (1961), who takes as the period he covers 1876 to 1957, when the *Journal of the Progressive Education Association* ceased publication two years after dissolution of the Association. Here we expand on some of the conceptual issues concerning the relationship between progressive education and psychoanalysis that were not a primary focus of his review.

The progressive education movement, whether we think of the broad span of activities reviewed in Cremin's book or the more formalized activities of the Progressive Education Association, which was founded in 1919, was not as concerned with the psychology of the child as with practical programs of school reform. It does not take a very substantial survey of the literature associated with the progressive education movement to determine that most of the interest was concerned with such matters as the organization of the schools, the content of the curriculum, and the innovative teaching methods that are now associated with the movement. Cremin (1961, pp. 306–308) gives a summary of the main impact of the progressive movement on the mainstream of American education, and of his ten summary points, none primarily concerns the psychology of the child. On the other hand, the move away from the traditional formal curriculum, the introduction of variation and flexibility in grouping of students, and the introduction of new curriculum materials are features of change which will be permanently associated with the movement.

Still, the progressive movement is popularly associated with the introduction of the concept of the child-centered school and in many cases

Sigmund Freud and other psychoanalysts were cited in support of this view, as in the following fragment of dialogue taken from Naumberg (1928):

If that is so, this new psychology would strike at the foundations of our present system of education.

It does, I think, just that. But nevertheless it will probably take another generation before the implications of psychoanalytic procedure begin to dislodge our present methods. (p. 205)

As another example, Agnes de Lima (1942), in the opening pages of her book *The Little Red School House*, acknowledges her indebtedness to "Dr. Sigmund Freud, who showed the tremendous role of the emotions in the education of the child" (p. 5).

Such examples are easily multiplied, but we do not want to give the impression that there was an extensive discussion of psychoanalytic ideas in the progressive literature. This was not the case, and the views that lay back of the introduction of the child-centered school, though they overlapped with many of the concepts of psychoanalysis, were by no means identical. To get a better sense of this, we summarize psychological points of emphasis in the doctrines of progressive education that were similar to those of psychoanalysis, as well as points that indicate separation between psychoanalysis and the central tendencies of progressive education. We begin with points of similarity.

1. The learning of each child should be considered on an individual basis and not committed to a uniform formal curriculum. The emotions of the child must be involved in his educational activities in order for learning to be successful.

2. The traditional tyrannical role of the teacher should be made a focus of change. The psychological character of this tyranny and the unconscious forces back of it should be explicitly recognized.

3. Traditional formal discipline in the classroom should be abolished and the child should be left as free as possible to do that which interests him. (The interpretation of what "free as possible" should mean varied enormously across different writers and schools, but the general idea remained one of the most important concepts of the child-centered school.)

4. A large number of learning difficulties can be solved by changing the traditional formal curriculum to an informal one centering around projects that engage the attention of the children.

The psychological points of disagreement are these.

1. That struggle and conflict are required in order for secondary processes to assume dominance over the more primitive primary processes was not seen as essential to learning, although from a psychoanalytic viewpoint the existence of this conflict is fundamental.

2. The limitations of what the teacher can accomplish and the inevita-

ble dangers of the power placed in teachers' hands, as emphasized by Erikson and Bernfeld, for example, received little if any recognition in the progressive movement. All that was needed, it was thought, was the reform of the teachers' attitudes so as to move from old to new methods of instruction. The fundamental psychological problems that exist in such relationships as surrogates for earlier family relationships between parents and child were scarcely recognized at all.

3. In the case of discipline there was certainly recognition by many progressive educators that the naive concept of absolute freedom for the child was not workable, but the deeper psychoanalytic insight that as the child matures the free instinctual forces must be repressed was scarcely recognized by any prominent progressive educators.

4. Concerning learning difficulties, the unconscious roots of such difficulties in some children were on occasion explicitly acknowledged by progressive educators, but their criticisms were so overwhelmingly directed at the formal methods of the past that the origin of learning difficulties in the early internal life of the child was seldom stressed.

The themes we have outlined were reflected in the pages of the two leading progressive journals. *The Teacher's College Record*, for example, from the first volume on, was most concerned with questions of curriculum and training. Volume 1 (1900) is subtitled "A Journal Devoted to the Practical Problems of Elementary and Secondary Education and the Professional Training of Teachers," and articles in this first volume include: "The Function of the University in the Training of Teachers" and "Syllabus for Teachers College Courses in Education."

There is little explicit impact of psychoanalysis to be found in the journal; in an article by Judd (1926) entitled "A Century of Applications of Psychology to Education," there is no mention of Freud at all. Indeed the first article wholly written from a psychoanalytic point of view is one by Goodwin Watson in 1957, who in an interesting way stresses the significance of the psychoanalytic approach to learning.

There is evidence, however, of an indirect impact of psychoanalysis, for example, in "Sex Education in Parental Education" by M. A. Bigelow in 1929, "The Relation of Family and School Life in the Education of Children" by L. H. Meek in 1934, and several articles by Symonds. Although there is little explicit reference to Freud or psychoanalysis, from about 1915 onward there is continual casual use of a psychological vocabulary that at least partly derives from psychoanalysis: *consciously, unconsciously, subconscious* (1915); *impulse, consciousness, instinctive* (1920); *ego, aggressive, self-consciousness* (1933); *unconsciously, unconscious, aggression, sublimation, fantasies* (1942); *conscious, ego, aggressive* (1952). These sample words are also found in nearly every year, but their relative frequency is constantly low.

The journal *Progressive Education*, which began publication in 1924, was

directed at teachers and parents rather than educational researchers. Although psychoanalytic concepts did not play a prominent role in any year of publication of the journal, there were occasional articles explicitly concerned with psychoanalysis and its relation to education.

One of the best of these articles is by Coriat (1926), who discusses the child's transference of emotion from parent to teacher, and the problems of understanding these unconscious emotional ties. He also stresses that "teachers must realize that the school forces the child to sublimate his instinctive libidinal energies" (p. 22), and that apparent learning failures must be carefully examined for evidence of unconscious resistance. In connection with this latter point he has a detailed discussion of fallacies of interpretation of a psychometric test of mental ability that arise from naive classification of students' oral responses.

Wittels (1931) emphasizes the usefulness of psychoanalysis in helping teachers relate to children, parents, and themselves. Taft (1932) analyzes the contributions of the psychoanalyst Otto Rank to education. Robison (1949, 1954) compares Freudian psychoanalysis and the work of neo-Freudians like Karen Horney. Kerlinger (1956) assesses the contributions of Freud and Dewey to attitudes of permissiveness in school discipline.

The sample we have briefly commented on is not complete, but it is representative and does constitute a reasonable proportion of the articles on psychoanalysis that appeared in *Progressive Education*.

3.5 Cultural Impact

A large number of historians of art, literature, and society have analyzed and commented on the remarkable influence Freud has had on American culture in the 20th century (e.g., Sullivan, 1932; Hofstadter, 1955; Hoffmar, 1957; Hale, 1971). It has been beyond our resources to analyze this larger impact in terms of the great wealth of primary materials in books, magazines, and newspapers of the time. What we have done is to collect a sample of quotations arranged chronologically so as to provide in a more vivid way than would a summary of the secondary historical literature mentioned above a sense of how Freud was viewed, especially by the educated public. Certainly many of the sources we cite are magazines that were read by only a very small fraction of the country's elementary-school teachers, but they are journals that are generally thought to be leaders in the formation of public opinion, or at least one important segment of such opinion.

To begin with, Walter Lippmann had this to say in *The New Republic* in 1915:

When I compare [Freud's] work with the psychology that I studied in college, or with most of the material that is used to controvert him, I cannot help feeling that for his illumination, for his steadiness and brilliancy of mind, he may rank among the greatest who have contributed to thought. I know how

easy it is to be deceived, but I take it that this is a small risk in comparison with the necessity for recognizing in his own lifetime a man of outstanding importance. After all, there were people who welcomed Darwin, and saw how profoundly he must affect our thinking. In Freud I believe we have a man of much the same quality, for the theories that have grown from his clinic in Vienna have always flowered in endless ways. From anthropology through education to social organization, from literary criticism to the studies of religions and philosophies, the effect of Freud is already felt. He has set up a reverberation in human thought and conduct of which few as yet dare to predict the consequences. (p. 10)

Two years later an unsigned article ("Is the Psycho-Analyzed Self a Libel on the Human Race?" 1917) in *Current Opinion* of a more negative character admits how familiar Freudian ideas have become:

Freud and Freudian interpretations have become familiar if contested issues. Whether the principles they posit express a keen insight into the frailties of the human, all too human, or constitute a libel upon its fair name is the tangible point of discussion, but we must not inject the moral issue into an investigation of purely scientific fact. If the Freudian interpretation of human motives and of their mechanisms and relation be sound, it is just as valuable whether its statement is pleasant or unpleasant, fit for family reading or only for the expert's library. (p. 105)

A year later, a review in *The New Republic* of Freud's *Totem and Taboo* begins this way ("Totem and Taboo," 1918):

It is now a good many years since William James made his first references to Freud, but it is only in the last ten years that Freud's name has come to be well circulated in America. (p. 347)

And further on in the same review:

In the end the knowledge of man's psyche may no more resemble Freud's than our present knowledge of the circulation of the blood resembles the Greeks'. But the adventure opened out by Freud is beyond doubt a supreme adventure. (p. 347)

In 1920, an unsigned article ("Is Freudism Destined to Live?") in *Current Opinion* on psychoanalysis begins:

Whatever value the thinking world may ultimately set on that method of treating nervous disorders originated by Dr. Sigmund Freud, of Vienna, and known as psychoanalysis, it is indisputable that, at the present time, the Freudian theories are showing extraordinary vitality. Freud has disciples in almost all the European countries, as well as in America. One can hardly pick up a newspaper or a magazine without finding psychoanalytic terms. The "conflicts" from which we all suffer; the "ambivalent" attitude which we feel toward so many people and things; the "suppressed desires" which poison our lives, are becoming subjects of daily conversation. (p. 355)

The next year an unsigned article ("A Physician's Warning Against Psycho-Analysis," 1921) in the same journal admits the currency of psychoanalysis and warns against its dangers:

As proof of his assertion that there is no kind of lunacy which will not be welcomed if it but masquerade in the garb of science, Doctor George Matheson Cullen cites psycho-analysis. Had this cult, he thinks, remained in a select professional circle, it might be passed by in silence. It has entered the marketplace and it has been popularized. Hence a warning must be issued against it. (p. 222)

A 1925 article in the *Forum* by George Sylvester Viereck opens with this splendid tribute:

Sigmund Freud has played an important part in the intellectual life of the world so long that, like Bernard Shaw, he has almost ceased to be a person. He is a cultural force to which we can assign a definite historical place in the evolution of civilization. (p. 302)

Later in the same article he continues:

In America, more than anywhere else, not only scientists but poets, novelists, and playwrights are mastering the laws of psycho-analysis. Harvey O'Higgins, Rupert Hughes, among our older novelists, freely admit their indebtedness to Freud in their more recent novels. Sherwood Anderson, O'Neill, Ben Hecht, James Branch Cabell are unthinkable without Freud. Surely Jurgen is a traveler through the realm of the unconscious! Edgar Lee Masters, in his *Spoon River Anthology*, indulges spasmodically in psycho-analysis. (p. 310)

Beginning in 1926, when Freud was seventy, a number of tributes testifying to his influence appeared, but the mere public notice of his various birthdays is in itself the strongest evidence of impact.

In 1926, *Living Age* said this ("Life, Letters, and the Arts"):

No name in modern psychology, not even that of William James, has become so much a part of the general vocabulary as that of Sigmund Freud, the great Viennese psychiatrist, who has recently celebrated his seventieth birthday. (p. 689)

In 1931, *The Nation* paid this tribute:

At the age of seventy-five Sigmund Freud has lived to find himself, with Albert Einstein, one of the two minds that have most profoundly influenced the thought of our time. It is true that the physicist's work seems much more securely established than the psychoanalyst's, and that many of Freud's theories are still the subject of heated controversy. But to one person who understands the theory of relativity or its implications there are a hundred whose outlook on life has been deeply colored by the theories of Freud. (*The Nation*, May 20, 1931, p. 545)

And again in 1936, *The Nation* said:

At eighty Sigmund Freud is, of all living men, the one who has most profoundly influenced the thinking of his contemporaries. Even Einstein is not a close rival. . . . If the whole of his system were to explode tomorrow, his influence would continue for centuries merely through the adoption into our language of his terminology and his metaphors—precisely as the influence of other exploded sciences, like alchemy and astrology and phrenology, has persisted. Not, of course, that there seems any present likelihood that it will be

exploded. The main features of his method and the main outlines of his conception have been quietly assimilated by the medical schools and sanatoria which were looking askance at him not more than twenty years ago. (*The Nation*, May 20, 1936, p. 631)

In 1937, George Soule has this to say at the beginning of an article in *The New Republic* on psychology and revolution:

If one were asked to name the two thinkers who have had the most influence on ideas and institutions in the time of persons now living, the answer would almost certainly be Karl Marx and Sigmund Freud. However critical one may be of the theory of either, it must be admitted that their conceptions have become an indispensable part of the apparatus of thought. Many of their word-labels live daily in the language of persons who are unconscious of the derivation of the phrases in question and never have understood fully what the authors meant by them. What should we do without "capitalism" or "complex"? (p. 66)

The comparison by Soule of Freud and Marx mirrors the earlier comparison above of Freud and Einstein.

To change the focus somewhat, here is a quotation from *Time* magazine in 1938 summarizing Thomas Mann's view of Freud as a literary figure:

Hailing Freud as a master of prose, Novelist Thomas Mann summed up the feeling of many a modern writer when he called him the greatest living man of letters, whose discoveries pointed the way to an art which might be bolder, freer, blither "than any possible in our neurotic, fear-ridden, hate-ridden world." (*Time*, May 23, 1938, p. 61)

The next quotation is again from *Time* in 1938 reporting on the sale of the Modern Library edition of Freud:

In the U.S. heyday of Freudian psychology during the 20s, nearly every intelligentsiac bought at least one simple popularization of Freud's works and could reel off an impromptu psychoanalysis at the drop of a symbol. With Depression, Freud was more & more often supplanted either by such former disciples as Alfred Adler, who called his adaptation "Individual Psychology," or by Karl Marx. To some observers, Freud's declining popularity among common readers looked permanent.

But three months ago *The Basic Writings of Sigmund Freud* was published as a Modern Library Giant (*Time*, May 23). The original edition, cautiously figured to last six months, was 10,000 copies. With few critics paying it attention, it sold out in eleven days. A second edition of 20,000 copies has now dwindled to its last few thousand copies, and Freud has come back as the fastest-selling reprint throughout the U.S. (*Time*, August 22, 1938, p. 55)

In 1939, Karl Menninger paid this moving tribute in *The Nation* to Freud, whose death had just occurred at age eighty-four:

No one in the field of psychology ever attained to a fraction of his stature. Among medical scientists almost none can be said to have approached him in brilliancy, originality, or influence upon medical practice. Perhaps no other one individual in the field of science lived to see the thinking of the entire

world so profoundly modified by his discoveries within his lifetime as did Freud. Galileo, Dalton, Lavoisier, Darwin, these and others contributed discoveries which greatly modified our thinking and our ways of living, but the effect was more gradual in its permeation. For not only medical science and psychological science and sociological science, but literature, art, anthropology, pedagogy, and even popular speech show the influence of Freud's discoveries and show them in unmistakable terms. (p. 373)

In 1947, Harry McNeill, writing in the Catholic journal *The Commonweal*, has this to say about the relation between psychoanalysis and Catholics:

If, however, the Freudians in their role of moral guides, as remolders of character, have much to learn from old hands at this game, Catholics too have much to derive from the Moses of modern psychiatry and his followers.

To begin with, Catholics must recognize the fact and the importance of the unconscious. This is really something that should require little prodding. Actually, we should not need a Freud to remind us of this when our traditional psychology is based upon four cardinal concepts: power, disposition, habit, and act, and of these four we can become directly conscious of act alone.

Having realized under the promptings of Freud the implications of our own doctrines, we should take the next step and recognize the instrument, par excellence, for investigating the unconscious, namely, psychoanalysis. (p. 351)

The next quotation from Bernard DeVoto in *Harper's Magazine* in 1948 comments on Freud's impact on art.

I suppose that in all history no other scientist has had so immediate or so profound an effect on art as Freud. And no other scientist was ever so fascinated by art or used it so often in the development of his science. Freud never systematized his ideas in a treatise on art as such and so they are scattered through his work with little order or sequence. They are usually expressed in regard to other things and usually as part of a highly technical mechanism of psychoanalysis. But anyone can find them and they seem to me in the main consistent, no more modified by the passage of years than the ideas of any scientist are modified as he accumulates more data and as his understanding of them deepens. They are not, however, simple ideas, they are exceedingly complex, for the method of exploration he was developing kept finding ever more complex processes beneath the simplest psychic experiences. (p. 97)

In 1952, W. H. Auden had a number of interesting things to say about Freud in *The New Republic*:

Today, thanks to Freud, the man-in-the-street knows (to quote by an inaccurate memory from *Punch*) that, when he thinks a thing, the thing he thinks is not the thing he thinks he thinks, but only the thing he thinks he thinks he thinks. Fifty years ago, a girl who sprained her ankle on the eve of a long-looked-forward-to ball, or a man who suffered from a shrewish wife, could be certain of the neighbors' sympathy; today the latter will probably decide that misfortune is their real pleasure. The letter of apology to the hostess whose dinner invitation you have forgotten is much more difficult to write than it used to be. If an Isolde worries all day lest her absent Tristan should be run

over by a bus, the dumbest Brangaene could warn her that her love includes a hope that he will never return.

Freud certainly expected opposition and obloquy from the conventional moralists and the man-in-the-street for his theories about human sexuality; in actual fact, the general public took him to their bosoms rather less critically, perhaps, than they should have done, while the real opposition came and still comes from the behaviorists, the neurologists and all the schools of psychiatry that regard their subject as a natural science. (pp. 16-17)

Finally, we cite Alfred Kazin writing in *The New York Times* in 1956. Kazin emphasizes both the impact of Freud and the controversial and speculative character of his writing. The quotation is from the opening paragraphs of the article.

It is hard to believe that Sigmund Freud was born 100 years ago today. Although Freud has long been a household name (and, in fact, dominates many a household one could mention), his theories still seem too "advanced," they touch too bluntly on the most intimate side of human relations, for us to picture Freud himself coming out of a world that in all other respects now seems so quaint.

Although Freud has influenced even people who have never heard of him, not all his theories have been accepted even by his most orthodox followers, while a great many of his essential ideas are rejected even by many psychoanalysts. In one sense Freud himself is still battling for recognition, for because of the tabooed nature of the materials in which he worked and the unusually speculative quality of his mind, Freud still seems to many people more an irritant than a classic.

On the other hand, Freud's influence, which started from the growing skepticism about civilization and morality after the First World War, is now beyond description. (p. 22)

The seventeen authors cited above have not been selected to be properly balanced between positive and negative attitudes toward psychoanalysis; most of the passages cited are favorable. What we were interested in exhibiting was the sense of general public awareness of Freudian theory. Even the early citations with strongly negative viewpoints (e.g., *Current Opinion* in 1917) acknowledge the widespread impact of Freud's ideas. Although we have attempted no systematic study by, for example, using an appropriate sampling design, it is our impression that for the latter part of the period we survey, i.e., from about 1930 to 1957, the preponderance of remarks about psychoanalysis in the more intellectually sophisticated public journals was positive in tone.

Parent education. Another source of change in thinking came through the parent-education movement. Though there seems to be conflicting evidence on the marital status of teachers during the period under review, it is unlikely that a movement that reached such a wide audience would have not reached, as part of this audience, teachers as well. During the period of our survey, parent education developed into a large mass movement; a good contemporary history of parent education is to be

found in Thomas and Thomas (1928).¹⁰

Stendler (1950) has analyzed three popular women's magazines, *Ladies Home Journal*, *Woman's Home Companion*, and *Good Housekeeping*, for the years 1890–1948, for advice to mothers regarding child rearing. Several trends emerge from her analysis. First is the dominant concern with physical development such as we noted in our discussion of the child-development texts. Second is the change in emphasis from character development to personality development as seen in the change from a rigid, disciplinary approach to child rearing (1910 through the 1930s) to an emphasis on self-regulation and understanding of the child (1940s). Stendler suggests that the reasons for these revolutionary changes must be looked for in general cultural changes rather than in "established" science, and she suggests that at least by the late 1930s much of the change can be attributed to the general influence of Freud.¹¹

We conclude with some remarks about the most important American child-rearing manual of the 20th century, namely, Benjamin Spock's *Baby and Child Care*, first published in 1946 and subsequently to go through almost 200 printings. Spock does not have strong psychoanalytic affinities and he does not often use explicit psychoanalytic terminology. His language is simple and is written in a commonsense vocabulary. On the other hand, Sulman (1973) has argued by the use of contextual content analysis that much of the orientation of *Baby and Child Care* is close in conceptual formulation and content to psychoanalytic thought. A similar view, based on an analysis of Sulman's results, is expressed by Bach (1974), who emphasizes that Spock generally accepts the psychoanalytic view of child development in terms of infantile sexuality, Oedipal conflict, and the active role of the unconscious, even though the more technical psychoanalytic concepts are not presented either in standard psychoanalytic terms or in commonsense surrogates.

A valid causal analysis would be difficult to give and certainly is beyond our resources, but on the basis of a comparison of Spock's famous manual with the most popular manual of the 1920s, Emmett Holt's *The Care and Feeding of Children* (originally published in 1894), it is reasonable to conjecture a general influence of psychoanalytic thought, because Holt's book

¹⁰ Of particular importance for the dissemination of psychoanalytic thinking was the Child Study Association of America. Its journal, *Child Study*, under the guiding editorship of S. M. Gruenberg, was explicitly psychoanalytic in orientation, and many of the articles were written by analysts themselves, ordinarily at least one article per volume. The articles, on the whole, were sophisticated in their method of presentation, and seemed to reflect ongoing changes in the analytic point of view, particularly in the stress on limits.

¹¹ Elsewhere, Steere (1968) has suggested that, contrary to what might be expected, psychoanalytic ideas did not have influence on the child-rearing manuals used up to about 1930. If Stendler's analysis is correct, it is probable that Steere did not continue his analysis far enough. Despite the interest in psychoanalysis by the popular press and the urban intelligentsia during the 1920s and 1930s, it seems to have taken about two decades for the ideas to become incorporated in the more conservative manuals and textbooks.

reflects a simple combination of behaviorism and medical advice. The contrast between the content of Holt's manual and Spock's could, without exaggeration, be characterized as primarily the absence of psychoanalytic influence in the former and its presence in the latter.

4. A COMPARISON OF FREUD, DEWEY, AND THORNDIKE

In this section we compare in general terms the differing conceptions of human nature to be found in the writings of Freud, Dewey, and Thorndike. Such a comparison sheds light on the differing philosophies that influenced the mainstream of elementary education during the period under review and shows how far apart different groups of educators and psychologists were in their theoretical approaches to human nature. The selection of Dewey and Thorndike was not based on a study of their impact in any depth, although some data on Thorndike are given in Table 3, but their large influence on American education in the first half of this century is widely accepted and does not require detailed documentation.

We begin by sketching Dewey's views, relying especially upon his influential book *Human Nature and Conduct* (1922), a three-part treatise devoted to habit, impulse, and intelligence.

Habits are defined by Dewey in Part I in the following passage:

We may think of habits as means, waiting, like tools in a box, to be used by conscious resolve. But they are something more than that. They are active means, means that project themselves, energetic and dominating ways of acting. (p. 25)

The place of habits in psychological and moral theory is outlined as follows:

Any moral theory which is seriously influenced by current psychological theory is bound to emphasize states of consciousness, an inner private life, at the expense of acts which have public meaning and which incorporate and exact social relationships. A psychology based upon habits (and instincts which become elements in habits as soon as they are acted upon) will on the contrary fix its attention upon the objective conditions in which habits are formed and operate. The rise at the present time of a clinical psychology which revolts at traditional and orthodox psychology is a symptom of ethical import. It is a protest against the futility, as a tool of understanding and dealing with human nature in the concrete, of the psychology of conscious sensations, images and ideas. It exhibits a sense for reality in its insistence upon the profound importance of unconscious forces in determining not only overt conduct but desire, judgment, belief, idealization. (p. 86)

Although the general tenor of this paragraph sounds sympathetic to psychoanalytic themes, so that one might expect to find in Part 2, on impulses, a theory compatible with psychoanalysis, this is, however, far

from the case. Rather, Dewey's continual emphasis on social questions and his continued animadversions against a highly individualistic psychology make his views quite different from psychoanalytic ones. The following quotation makes clear the importance for him of the relation of impulses to the variety and character of social institutions:

If we inquire why men were so long blind to the existence of powerful and varied instincts in human beings, the answer seems to be found in the lack of a conception of orderly progress. It is fast becoming incredible that psychologists disputed as to whether they should choose between innate ideas and an empty, passive, wax-like mind. For it seems as if a glance at a child would have revealed that the truth lay in neither doctrine, so obvious is the surging of specific native activities. But this obtuseness to facts was evidence of lack of interest in what could be done with impulses, due, in turn, to lack of interest in modifying existing institutions. It is no accident that men became interested in the psychology of savages and babies when they became interested in doing away with old institutions.

A combination of traditional individualism with the recent interest in progress explains why the discovery of the scope and force of instincts has led many psychologists to think of them as the fountain head of all conduct, as occupying a place before instead of after that of habits. (pp. 93-94)

Perhaps more startling is this passage describing the role of impulse in the psychic life of the child or adult:

Impulse in short brings with itself the possibility but not the assurance of a steady reorganization of habits to meet new elements in new situations. The moral problem in child and adult alike as regards impulse and instinct is to utilize them for formation of new habits, or what is the same thing, the modification of an old habit so that it may be adequately serviceable under novel conditions. (p. 104)

A few pages later, Dewey says, in comparing custom and instinct, that custom is immutable but that "instincts are most readily modifiable through use, most subject to educative direction" (p. 107). From these and other passages it is clear that Dewey's theory of the instincts is very different from Freud's. The power of the sexual and aggressive instincts in the Freudian sense is not something that is admitted or recognized by Dewey.

In the third part, when Dewey turns to intelligence, he does bring out a theme that is closer to Freud than some of those mentioned. For both Dewey and Freud, activity does not result in an end perfectly attained, and both reject the concept of an Aristotelian end that is ultimate and final as a mistaken ideal, contrary to a widely held classical view of education. The difference between Freud and Dewey on this point is that Freud insists on the ineradicable character of conflict whereas Dewey minimizes the existence of internal conflict and characterizes activity as a means of solving problems that arise in the environment.

In sum, Dewey exhibits an ambivalent attitude toward psychoanalysis. For example, he objects to the psychoanalytic view that neurotic attitudes

and habits can be changed by purely psychological analysis without reference to sensation and perception (p. 34). He objects to the idea of a separate psychic realm and the existence of psychic forces or psychic energy (p. 87) and he is not happy with the kind of elaborate explanations of dreams and parapraxes given by Freud (p. 252). He objects (p. 133) even more strongly to basing so much of the theory on the sexual instinct and in general is not sympathetic to the prominent place given to sexual drives in psychoanalysis (p. 153). Throughout Dewey's many writings on education there seems to be hardly any place that he deals at any length with sexuality in children. He does mention sexual desires and sexual instincts in various passages but scarcely anywhere their presence in children. Above all, he does not assign to them the ubiquitous and significant role that we now do. The sexual openness of American society in the latter part of the 20th century and the increasing emphasis on an ever explicit sexual education of the young are a reflection of the deep influence of Freud on our conception of education and society rather than of a current of thought that was at all salient in Dewey.

It is no accident that Dewey begins the treatise under discussion with a lengthy analysis of habit, because, for him, habits have a primacy over either reason or instinct in the characterization of human nature. To some extent this emphasis on habit is due to the social focus of his psychology, and, in fact, Dewey subtitles his treatise *An Introduction to Social Psychology*. Still, we cannot differentiate the thought of Freud and Dewey only in this fashion, for it is characteristic of Dewey not only in *Human Nature and Conduct* but in many other places in his writings to inveigh against a psychology that is highly individualistic. His choice of subtitle reflects not simply a choice of specific topics but a fundamental view of the nature of psychology and of the essentially social character of human nature.

In the same vein, there is a striking difference between the view of human nature expressed by Dewey, not only in *Human Nature and Conduct* but in his many writings on education and democracy as well, and the more pessimistic view expressed by Freud in many places but especially in his essay *Civilization and Its Discontents* (1930/1961). In an important footnote in this work (p. 134) Freud emphasizes the importance in the education of the young of preparing them for the acts of aggression they will meet as they move out into the world. Freud's belief that aggression cannot be eliminated, and that the development of civilization is limited by it, stands in contrast to Dewey's more optimistic view.¹²

Edward L. Thorndike occupies a position in American education in the first half of this century that is related to but quite different from that of either Dewey or Freud. For one, he is known principally to a relatively

¹² Some more extensive comparisons of Freud and Dewey are to be found in Feuer (1960), Levitt (1960), and Karier (1963).

small group of specialists in educational psychology and curriculum, whereas Freud's name is familiar in every intellectual circle and Dewey has had extremely wide currency both in education and in philosophy. All the same, during the period covered by our review, Thorndike's influence in educational psychology and school curriculum was greater than that of either Dewey or Freud; educational psychology during this period was dominated by learning theory, and the quintessential figure in learning theory was Thorndike.

As in the case of Freud and Dewey, we do not attempt to survey systematically the voluminous writings of Thorndike in this summary comparison but concentrate on his devotion to learning theory and its applications, epitomized in the second volume of his monumental treatise, *Educational Psychology* (1913-1914). In this work, Thorndike's scientific objective was to show how essentially all of school learning could in principle be accounted for in terms of his three fundamental laws of readiness, exercise, and effect. The Law of Readiness stresses that the connection between stimulus and response needs to be satisfying to an organism. The Law of Exercise would be formulated today in terms of the strength of a response being increased by repeated trials and being decreased by the absence of practice. The most famous of the three laws, the Law of Effect, states that when a modifiable connection between a stimulus and a response is accompanied by a satisfying state of affairs or, as we might put it now, a positive reinforcement, then the connection is strengthened, and if it is accompanied by a negative reinforcement the connection is weakened.

One of the most impressive things about Thorndike is the strenuous scientific effort he made to bring these laws into the arena of significant educational application. He published two extensive empirical works on mathematics learning, one *The Psychology of Arithmetic* (1922) and the other *The Psychology of Algebra* (1923), and numerous additional research papers on reading and mathematics skills, as well as on many other topics that are part of the school curriculum. His extensive work on reading is reviewed in Geraldine Joncich Clifford's chapter in this volume. It is on the basis of this large corpus of work that Thorndike had his unrivaled influence on school curriculum and educational psychology in the period under review.

Thorndike's more general views about human nature are probably well conveyed in his long treatise *Human Nature and the Social Order* (1940). This work does not have the intellectual distinction of the similar works by Freud or Dewey, but it does have the virtue of attempting to treat a large number of highly specific features of human behavior that are of interest both to educators and to the public at large, ranging from the basis of the desire to smoke tobacco to the philosophical problem of concern for the

welfare of unborn generations. Although Thorndike uses many general psychological terms also used by Freud, e.g., wants, gratification, aversions, desires, and conflict, his actual theory can be most easily characterized as a kind of psychological utilitarianism. He shows a strong interest in developing an elementary quantitative theory of wants as reflected in utilities and disutilities and to build an economic model of much everyday behavior. Though this interest is laudatory, his use of economic concepts is superficial, and the work ultimately rests on principles of learning theory and theories of individual differences, especially as reflected in the theory of testing. Surprisingly enough, there is only one mention of Dewey and several of Freud, although Dewey and Thorndike were colleagues at Teachers College for many years.

The features of Freud's work that we have stressed as having been of great importance for education and for 20th-century views of human nature are almost entirely omitted by Thorndike. He has little to say about the unconscious; his theory of conflict scarcely touches the deeper reaches that concern Freud. The role of the defense mechanisms in psychological life is scarcely referred to in Freudian terms or as discussed by the literature we have reviewed. And finally and above all, the recognition of the important place of sexuality in childhood development receives scarcely any mention at all.

In concluding this comparison, there is one important respect in which Freud and Thorndike have more in common with each other than either does with Dewey. This is in the continual concern they both show for empirical data, Freud by means of clinical methods and case studies, Thorndike by laboratory experiments and by controlled methods of testing for individual differences in mental abilities. In contrast, Dewey, in spite of his general endorsement of scientific methods and naturalistic philosophy, did not, except at the very beginning of his career, engage in empirical inquiry, and the bulk of his writings are remarkable for their absence of even secondhand use of specific data. Although Freud and Thorndike are poles apart in general philosophy, this methodological similarity should not go unnoticed.

5. CONCLUSION

In the preceding pages we have attempted to present the part of psychoanalytic theory that is most relevant to the education of normal children and to assess its impact on aspects of American elementary education. In this final section we summarize our general conclusions. These conclusions overlap somewhat, but we discuss each one separately in order to bring out what we believe are the major findings of our inquiry.

5.1 *The Diffuse Influence of Psychoanalysis on Education*

Psychoanalysis has had a significant effect on 20th-century elementary education. Many psychoanalytic ideas are now so much a part of our thinking about education that it is hard to realize how much the climate has changed from a century ago. The most salient psychoanalytic ideas that were incorporated into educational thought were the importance of the early years of childhood, the fact of childhood sexuality, and the role the defense mechanisms play in internal life.

Although it seems reasonable to attribute to psychoanalysis much of the changed approach to young children in school, its explicit influence in the professional educational literature during the first half of this century was not substantial. Only the mental hygiene literature showed a substantial explicit influence of psychoanalytic ideas, but this literature in itself reflects a profound change in attitudes toward education. No longer are the central aims of education solely the acquisition of basic skills and a fund of knowledge but now include the development of a healthy and effective personality.

The indirect and derived impact of psychoanalysis was substantial and is reflected in the changing themes and emphases in texts, in the revised attitudes of textbook authors, in the treatment of topics like projective tests, and in the acceptance of childhood sexuality. Psychoanalysis developed in a medical setting and its history is evident in the main lines of its influence in education. The most easily accessible professional area was that of mental hygiene, with its concern for problem behavior, and this was indeed the area most affected. The specific contributions that psychoanalysis made to the more general psychology of child development and educational psychology were limited.

5.2 *The Increasing Influence of Psychoanalysis over Time*

In broad terms, the influence of psychoanalysis in education grew throughout the entire first half of this century, the period that we have surveyed. The fact that psychoanalytic ideas were incorporated more and more over this period is in its way evidence of the magnitude and depth of their impact. It is also evidence of how slow the application of new scientific ideas can be. The gradual growth of the influence of psychoanalysis stands in sharp contrast, for example, to the impact of some of the other scientific discoveries made at the end of the 19th century. Perhaps the most extreme example is Roentgen's discovery of X-rays in 1895. With no applications of any kind in mind and in fact by fortuitous accident, Roentgen, while studying the discharge of electricity through rarefied gases, discovered the phenomenon of X-rays. Within three months of his discovery, X-rays were being put to practical use in a

hospital in Vienna in the diagnosis and analysis of bone fractures.

Freud certainly had no educational applications in mind when he began his own early work, yet from the standpoint of our culture, his contributions have had a much more profound influence than most of the particular scientific discoveries that can be mentioned. However, it took a long and complicated route for the impact of psychoanalysis to be felt and for the ideas to become a part of our thinking about the child and therefore about the education of children.

5.3 *Ambivalence Toward Psychoanalysis*

The ambivalence toward psychoanalysis was exhibited in a number of different ways. Textbook authors we have cited from various fields of education more often than not adopted such an attitude in their explicit comments on psychoanalysis, and in the extensive presentations of the defense mechanisms it was uncommon for their origin in psychoanalytic thought to be given proper attribution. Although the almost uniform theoretical eclecticism of the textbook authors reflected an ambivalence toward theory in general, this was especially the case toward a theory whose ideas and methods ran contrary to the prevailing academic views and procedures.

Another reflection of ambivalence toward psychoanalysis is that the substantial developments in ego psychology did not have the impact on education that might have been expected. Although the explanation for this could be that the natural movement of ideas from the technical literature of a subject to the literature dealing with its applications is usually slow, or that the way in which ego psychology was developed by Hartmann (1939/1958) and others—especially the concept of the autonomous functions of the ego—did not make adequate contact with those developed parts of educational psychology dealing with learning and related matters, it is also no doubt the case that limited acquaintance with the psychoanalytic literature reflected limited interest.

5.4 *Nonprofessional Channels*

To a very considerable extent the channels through which the influence of psychoanalysis spread were not professional ones. It seems fair to say that the dissemination of psychoanalytic ideas through the public press and through the very large literature addressed to parents had as much influence as the professional literature in education itself. Teachers' attitudes changed on a variety of topics in a direction more consistent with psychoanalytic concepts over most of the period we have surveyed, and it seems likely that much of the basis for this change was the dissemination of psychoanalytic ideas throughout the general culture rather than through any particular professional literature.

5.5 *Difficulty of Causal Analysis*

The issues raised in the preceding paragraph about the sources of influence on teachers' attitudes force us to make explicit our recognition of the difficulty of providing an adequate analysis of these matters. A scientific theory as complicated as psychoanalysis follows a path of influence and neglect that makes the sharp identification of causal relations an extremely difficult matter. The changes in teachers' attitudes in a direction more favorable to the basic ideas of psychoanalysis are in our judgment to a significant degree attributable to the broad impact of psychoanalytic ideas on 20th-century American culture, but we are not able to give the kind of detailed causal analysis required to make this generalization scientifically unimpeachable. For practically every one of the general ideas mentioned at the beginning of this section, a case could be made for its having arisen elsewhere and being due to other influences than that of psychoanalysis. Nevertheless, the weight of the evidence of psychoanalytic influence, though inconclusive, is still substantial.

5.6 *Limitations of Education*

In contrast to an emphasis on unlimited growth and the unbounded potentialities of education characteristic of the writings of John Dewey and, more generally, of the progressive education movement, psychoanalytic thinking about education, whether in the home or in the school, has stressed the natural limitations of what can be achieved and what should be expected. The view of limitations ranges from accepting the necessary existence of conflict in the developing child to recognizing that teachers are subject to the same kinds of psychological problems in dealing with students that parents are in dealing with their children. This strong sense of the limitations of what education can hope to achieve was not a psychoanalytic theme that was widely adopted in the period we have studied, although it is a salient feature of the intellectual climate today.

REFERENCES

- Almy, M. *Child development*. New York: Holt, 1955.
- Anderson, J. D. Child development: An historical perspective. *Child Development*, 1956, 27, 181-196.
- Auden, W. H. Sigmund Freud. *New Republic*, October 6, 1952, pp. 16-17.
- Averill, L. A. *Psychology for normal schools*. Boston: Houghton Mifflin, 1921.
- Averill, L. A. *The hygiene of instruction*. Boston: Houghton Mifflin, 1928.
- Averill, L. A. *Mental hygiene for the classroom teacher*. New York: Pitman, 1939.
- Averill, L. A. *The psychology of the elementary-school child*. New York: Longmans, Green, 1949.
- Bach, W. G. The influence of psychoanalytic thought on Benjamin Spock's *Baby and Child Care*. *Journal of the History of the Behavioral Sciences*, 1974, 10(1), 91-94.

- Beers, C. H. *A mind that found itself: An autobiography*. New York: Longmans & Green, 1908.
- Benson, C. E., & Alteneider, L. E. Mental hygiene in teacher-training institutions in the United States: A survey. *Mental Hygiene*, 1931, **15**, 224-241.
- Bernard, H. W. College mental hygiene: A decade of growth. *Mental Hygiene*, 1940, **24**, 413-419.
- Bernard, H. W. *Mental hygiene for classroom teachers*. New York: McGraw-Hill, 1952.
- Bernfeld, S. B. *Sisyphus or the limits of education* (F. Lilge, trans.). Berkeley: University of California Press, 1973. (Originally published, 1925.)
- Bernstein, I. D. Psychoanalytic extensions of the S-R formula. *Psychoanalytic Review*, 1935, **22**, 158-168.
- Bettelheim, B. The social-studies teacher and the emotional needs of adolescents. *School Review*, 1948, **56**, 585-592.
- Bettelheim, B. *The children of the dream*. New York: Macmillan, 1969.
- Bigelow, M. A. Sex-education in parental education. *Teachers College Record*, 1929-1930, **31**, 522-527.
- Blair, G. M. Educational psychology, its development and present status. *University of Illinois Bulletin*, 1948, **46**(13). (Monograph)
- Blanchard, P. Psychoanalytic contributions to the problems of reading disabilities. *Psychoanalytic Study of the Child*, 1946, **2**, 163-187.
- Boynnton, P. L. *Psychology of child development*. Minneapolis, Minn.: Educational Publishers, 1938.
- Breckenridge, M. E., & Vincent, E. L. *Child development: Physical and psychological growth through the school years* (2nd ed.). Philadelphia: Saunders, 1949.
- Breitwieser, J. V. *Psychological education*. New York: Knopf, 1926.
- Brooks, F. D., & Shaffer, L. F. *Child psychology*. Boston: Houghton Mifflin, 1937.
- Bruce, W. F. The relations of educational psychology with general psychology. *Journal of Educational Psychology*, 1949, **40**, 261-266.
- Burnham, J. C. *Psychoanalysis and American medicine, 1894-1914: Medicine, science, and culture*. New York: International Universities Press, 1967.
- Burnham, W. H. A survey of the teaching of mental hygiene in the normal schools. *Mental Hygiene*, 1921, **5**, 19-45.
- Burnham, W. H. *The normal mind*. New York: Appleton, 1924.
- Carroll, H. A. *Mental hygiene: The dynamics of adjustment*. New York: Prentice-Hall, 1947.
- Carroll, H. A. *Mental hygiene: The dynamics of adjustment* (2nd ed.). New York: Prentice-Hall, 1951.
- Carroll, H. A. *Mental hygiene: The dynamics of adjustment* (3rd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1956.
- Chase, H. W. Psychoanalysis and the unconscious. *Pedagogical Seminary*, 1910, **17**, 281-327.
- Cole, L. E., & Bruce, W. F. *Educational psychology*. New York: World Book, 1950.
- Commins, W. D. *Principles of educational psychology*. New York: Ronald Press, 1937.
- Commins, W. D., & Fagin, B. *Principles of educational psychology* (2nd ed.). New York: Ronald Press, 1954.
- Coriat, I. H. The psycho-analytic approach to education. *Progressive Education*, 1926, **3**(1), 19-25.
- Cremin, L. A. *The transformation of the school: Progressivism in American education, 1876-1957*. New York: Knopf, 1961.
- Cronbach, L. J. *Educational psychology*. New York: Harcourt Brace, 1954.
- Cuff, N. B. What should be included in educational psychology? *Journal of Educational Psychology*, 1935, **26**, 680-688.

- Cuff, N. B. *Educational psychology*. Louisville, Ky.: Standard Printing, 1936.
- Cutts, N. E., & Moseley, N. *Practical school discipline and mental hygiene*. Boston: Houghton Mifflin, 1941.
- De Lima, A. *The little red school house*. New York: Macmillan, 1942.
- De Voto, B. The easy chair. *Harper's Magazine*, 1948, **197**, 97-100.
- Dewey, J. *Human nature and conduct*. New York: Holt, 1922.
- Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- Douglas, O. B. The present status of the introductory course in educational psychology in American institutions of learning. *Journal of Educational Psychology*, 1925, **16**, 396-408.
- Dutton, W. H. *The child-study movement in America from its origin (1880) to the organization of the Progressive Education Association (1920)*. Unpublished doctoral dissertation, Stanford University, 1945.
- Ekstein, R., & Motto, R. L. *From learning for love to love of learning*. New York: Reiss-Davis Child Study Center, 1969.
- Ellis, D. B., & Miller, L. W. Teachers' attitudes and child behavior problems. *Journal of Educational Psychology*, 1936, **27**, 501-511.
- Ellis, R. S. *Educational psychology*. Toronto: Van Nostrand, 1951.
- Emme, E. E. Content analysis of nineteen textbooks in educational psychology. *Education*, 1942, **63**, 168-171.
- Erikson, E. *Childhood and society*. New York: Norton, 1950.
- Falick, M. L., Peters, M., Leviit, M., & Rubenstein, B. G. Observations on the psychological education of teachers in a school-based mental-hygiene program. *Mental Hygiene*, 1954, **38**, 374-387.
- Fenichel, O. The means of education. *Psychoanalytic Study of the Child*, 1945, **1**, 281-292.
- Fenton, N. *Mental hygiene in school practice*. Stanford, Calif.: Stanford University Press, 1943.
- Feuer, L. S. The standpoints of Dewey and Freud: A contrast and analysis. *Journal of Individual Psychology*, 1960, **16**, 119-136.
- Freud, A. *Psychoanalysis for teachers and parents* (B. Low, trans.). Boston: Emerson Books, 1935.
- Freud, A. *The ego and the mechanisms of defense*. New York: International Universities Press, 1946. (Originally published, 1936.)
- Freud, A. Answering teachers' questions. In *The writings of Anna Freud: Indications for child analysis and other papers* (Vol. 4). New York: International Universities Press, 1968. (Originally published, 1952.)
- Freud, S. Inhibitions, symptoms and anxiety. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 20). London: Hogarth, 1959. (Originally published, 1926.)
- Freud, S. The ego and the id. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 19). London: Hogarth, 1961. (Originally published, 1923.)
- Freud, S. Civilization and its discontents. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 21). London: Hogarth, 1961. (Originally published, 1930.)
- Garrison, K. C., & Gray, S. *Educational psychology*. New York: Appleton-Century-Crofts, 1955.
- Gates, A. I. *Psychology for students of education*. New York: Macmillan, 1923.
- Gates, A. I. *Psychology for students of education* (Rev. ed.). New York: Macmillan, 1930.

- Gates, A. I., Jersild, A. T., McConnell, T. R., & Challman, R. C. *Educational psychology* (3rd ed.). New York: Macmillan, 1948.
- Gates, A. I., Jersild, A. T., McKillop, A. S., Rivlin, H. N., Shoben, E. J., Jr., & Watson, G. Educational psychology. *Review of Educational Research*, 1956, **26**(3), 241-267.
- Gesell, A. L., & Gesell, B. C. *The normal child and primary education*. Boston: Ginn, 1912.
- Gesell, A., & Ilg, F. L. *The child from five to ten*. New York: Harper, 1946.
- Goodenough, F. L., & Anderson, J. E. *Experimental child study*. New York: Century, 1931.
- Groves, E. R., & Blanchard, P. *Introduction to mental hygiene*. New York: Holt, 1930.
- Guthrie, E. R., & Powers, F. F. *Educational psychology*. New York: Ronald Press, 1950.
- Hale, N. G., Jr. *Freud and the Americans: The beginnings of psychoanalysis in the United States, 1876-1917*. New York: Oxford University Press, 1971.
- Hall, G. S. *Contents of children's minds*. New York: Kellogg, 1893.
- Hall, G. S. *Educational Problems* (2 vols.). New York: Appleton, 1911.
- Hall, G. S. *Aspects of child life and education*. New York: Appleton, 1921.
- Hall-Quest, A. L. Present tendencies in educational psychology. *Journal of Educational Psychology*, 1915, **6**, 601-614.
- Harris, H. I. Repression as a factor in learning theory. *Psychoanalytic Quarterly*, 1950, **19**, 410-411.
- Hartmann, H. Notes on the theory of sublimation. *Psychoanalytic Study of the Child*, 1955, **10**, 9-29.
- Hartmann, H. *Ego psychology and the problem of adaptation*. New York: International Universities Press, 1958. (Originally published, 1939.)
- Hartmann, H., Kris, E., & Loewenstein, R. M. Comments on the formation of psychic structure. *Psychoanalytic Study of the Child*, 1946, **2**, 11-38.
- Hartmann, H., & Loewenstein, R. M. Notes on the superego. *Psychoanalytic Study of the Child*, 1962, **17**, 42-81.
- Healy, W. *The individual delinquent*. Boston: Little, Brown, 1915.
- Heidbreder, E. *Seven psychologies*. New York: Appleton-Century-Crofts, 1933.
- Hertzberg, O. E. The opinion of a teacher-training institution concerning the relative value of subject-matter in educational psychology to the elementary school teacher. *Journal of Educational Psychology*, 1928, **19**, 329-342.
- Hilgard, E. R. Human motives and the concept of self. *American Psychologist*, 1949, **4**, 374-382.
- Hilgard, E. R. *Theories of learning* (2nd ed.). New York: Appleton-Century-Crofts, 1956. (Originally published, 1948.)
- Hill, J. C. Poetry and the unconscious. *British Journal of Psychology, Medical Section*, 1924, **4**(2), 125-133.
- Hill, J. C. *Teaching and the unconscious mind*. New York: International Universities Press, 1971.
- Jobson, C. V. How much do teachers know about mental hygiene? *Mental Hygiene*, 1937, **21**, 231-242.
- Hoffer, W. Psychoanalytic education. *Psychoanalytic Study of the Child*, 1945, **1**, 293-307.
- Hoffman, F. J. *Freudianism and the literary mind* (2nd ed.). New York: Grove Press, 1957.
- Hofstadter, R. *The age of reform: From Bryan to F.D.R.* New York: Knopf, 1955.
- Hollingsworth, H. L. *Educational psychology*. New York: Appleton-Century, 1933.

- Holt, E. *The care and feeding of children*. New York: Appleton, 1894.
- Homburger, E. Psychoanalysis and the future of education. *Psychoanalytic Quarterly*, 1935, **4**, 50-68.
- Horne, H. H. *The psychological principles of education*. New York: Macmillan, 1906.
- Hunter, E. C. Changes in teachers' attitudes toward children's behavior over the last thirty years. *Mental Hygiene*, 1957, **41**, 3-11.
- Hurlock, E. B. *Modern ways with children*. New York: McGraw-Hill, 1943.
- Hurlock, E. B. *Modern ways with children* (2nd ed.). New York: McGraw-Hill, 1950.
- Hurlock, E. B. *Modern ways with children* (3rd ed.). New York: McGraw-Hill, 1956.
- Is Freudism destined to live? *Current Opinion*, 1920, **69**, 355-358.
- Is the psycho-analyzed self a libel on the human race? *Current Opinion*, 1917, **63**, 105.
- James, W. Abstract of Freud. *Psychological Review*, 1894-1895, **1**, 199.
- James, W. *Talks to teachers*. New York: Norton, 1958.
- Jensen, A. S. *Psychology of child behavior*. New York: Prentice-Hall, 1938.
- Jersild, A. T. *Child psychology* (3rd ed.). New York: Prentice-Hall, 1947.
- Jersild, A. T. Child psychology in the United States. *Teachers College Record*, 1948-1949, **50**, 114-127.
- Jersild, A. T. *Child psychology* (4th ed.). New York: Prentice-Hall, 1954.
- Jordan, A. M. *Educational psychology*. New York: Holt, 1928.
- Jordan, A. M. *Educational psychology* (Rev. ed.). New York: Holt, 1933.
- Jordan, A. M. *Educational psychology* (3rd ed.). New York: Holt, 1942.
- Jordan, A. M. *Educational psychology* (4th ed.). New York: Holt, 1956.
- Judd, C. H. A century of applications of psychology to education. *Teachers College Record*, 1926, **27**, 771-781.
- Judd, C. H. *Educational psychology*. Boston: Houghton Mifflin, 1939.
- Karier, C. J. The rebel and the revolutionary: Sigmund Freud and John Dewey. *Teachers College Record*, 1963, **64**, 605-613.
- Kazin, A. The Freudian revolution analyzed. *New York Times Magazine*, May 6, 1956, pp. 22, 37-38.
- Kerlinger, F. N. The origin of the doctrine of permissiveness in American education. *Progressive Education*, 1956, **33**, 161-165.
- Kirkpatrick, E. A. *Fundamentals of child study* (Rev. ed.). New York: Macmillan, 1907.
- Kirkpatrick, E. A. *The individual in the making*. Boston: Houghton Mifflin, 1911.
- Kirkpatrick, E. A. *Fundamentals of child study* (4th ed.). New York: Macmillan, 1929.
- Klein, E. Psychoanalytic aspects of school problems. *Psychoanalytic Study of the Child*, 1949, **3-4**, 369-391.
- Klein, M. The role of the school in the libidinal development of the child. *International Journal of Psycho-Analysis*, 1924, **5**, 312-331.
- Kris, E. On psychoanalysis and education. *American Journal of Orthopsychiatry*, 1948, **18**, 622-635.
- Kris, E. Notes on the development and on some current problems of psychoanalytic child psychology. *Psychoanalytic Study of the Child*, 1950, **5**, 24-46.
- La Rue, D. W. *Psychology for teachers*. New York: American Book Company, 1920.
- La Rue, D. W. *Mental hygiene*. New York: Macmillan, 1927.
- Lashley, K. S. Contributions of Freudism to psychology. III. Physiological analysis of the libido. *Psychological Review*, 1924, **31**, 192-202.
- Leviitt, M. *Freud and Dewey on the nature of man*. New York: Philosophical Library, 1960.
- Levy, D. M. *Maternal overprotection*. New York: Columbia University Press, 1943.

- Life, letters, and the arts: Freud at seventy. *Living Age*, 1926, **329**, 689.
- Lindgren, H. C. *Mental hygiene in education*. New York: Holt, 1954.
- Lippmann, W. Freud and the layman. *New Republic*, April 17, 1915, pp. 9-10.
- Liss, E. Libidinal fixations as pedagogic determinants. *American Journal of Orthopsychiatry*, 1935, **5**, 126-131.
- Low, B. Review of Hall's *Aspects of child life and education*. *International Journal of Psychoanalysis*, 1923, **4**, 211-214.
- Low, B. The teacher's problems—A psycho-analytical assessment. *New Era*, 1938, **19**, 227-229.
- Magnusson, P. M. *Psychology as applied to education*. Boston: Silver, Burdett, 1913.
- Malmquist, C. P. Conscience development. *Psychoanalytic Study of the Child*, 1968, **23**, 301-331.
- Martin, W. E., & Stendler, C. B. *Child development: The process of growing up in society*. New York: Harcourt, Brace, 1953.
- McNeill, H. Freudians and Catholics. *Commonweal*, 1947, **46**, 350-353.
- Meek, L. H. The relation of family and school life in the education of children. *Teachers College Record*, 1934-1935, **36**, 271-278.
- Menninger, K. *The human mind*. New York: Crofts, 1931.
- Menninger, K. Sigmund Freud. *The Nation*, 1939, **149**, 373-374.
- Millard, C. V. *Child growth and development in the elementary school years*. Boston: Heath, 1951.
- Mitchell, J. C. A study of teachers' and mental hygienists' ratings of certain behavior problems of children. *Journal of Educational Research*, 1942, **36**, 292-307.
- Morgan, J. J. B. *The psychology of the unadjusted school child*. New York: Macmillan, 1924.
- Morgan, J. J. B. *Child psychology* (3rd ed.). New York: Farrar & Rinehart, 1942.
- Munsterberg, H. *Psychology and the teacher*. New York: Appleton, 1909.
- Mursell, J. L. *Educational psychology*. New York: Norton, 1939.
- Mursell, J. L. *Psychology for modern education*. New York: Norton, 1952.
- Musser, P. H., & Conger, J. J. *Child development and personality*. New York: Harper, 1956.
- Nagge, J. W. *Psychology of the child: Mental and physical growth*. New York: Ronald Press, 1942.
- Naumberg, M. *The child and the world*. New York: Harcourt, Brace, 1928.
- Norsworthy, N., & Whitley, M. T. *The psychology of childhood*. New York: Macmillan, 1918.
- Oberndorf, C. P. *A history of psychoanalysis in America*. New York: Grune & Stratton, 1953.
- O'Shea, M. V. *Dynamic factors in education*. New York: Macmillan, 1906.
- O'Shea, M. V. Psychology in the normal school. *Journal of Educational Psychology*, 1911, **2**, 327-333.
- O'Shea, M. V. *Mental development and education*. New York: Macmillan, 1921.
- Park, D. G. Freudian influence on academic psychology. *Psychological Review*, 1931, **38**, 73-85.
- Partridge, G. E. *An outline of individual study*. New York: Sturgis & Walton, 1910.
- Pearson, G. H. J. A survey of learning difficulties in children. *Psychoanalytic Study of the Child*, 1952, **7**, 322-386.
- Pechstein, L. A., & Jenkins, F. *Psychology of the kindergarten-primary child*. Boston: Houghton Mifflin, 1927.
- Peck, P. Teachers' reports of the problems of unadjusted school children. *Journal of Educational Psychology*, 1935, **26**, 123-138.

- Peller, L. E. Incentives to development and means of early education. *Psychoanalytic Study of the Child*, 1946, **2**, 397-415.
- Peller, L. E. The school's role in promoting sublimation. *Psychoanalytic Study of the Child*, 1956, **11**, 437-449.
- Peterson, H. A. *Educational psychology*. New York: Macmillan, 1948.
- A physician's warning against psycho-analysis. *Current Opinion*, 1921, **71**, 222-224.
- Pintner, R. *Educational psychology*. New York: Holt, 1929.
- Plank, E. N., & Plank, R. Emotional components in arithmetical learning as seen through autobiographies. *Psychoanalytic Study of the Child*, 1954, **9**, 274-293.
- Pressey, S. L. *Psychology and the new education*. New York: Harper, 1933.
- Pressey, S. L., & Robinson, F. P. *Psychology and the new education* (Rev. ed.). New York: Harper, 1944.
- Pvor, H. B. *As the child grows*. New York: Silver, Burdett, 1943.
- Pyle, W. H. *The outline of educational psychology*. Baltimore: Warwick & York, 1911.
- Keddl, F., & Wattenberg, W. W. *Mental hygiene in teaching*. New York: Harcourt Brace, 1951.
- Reynolds, M. M. *Children from seed to saplings*. New York: McGraw-Hill, 1939.
- Reynolds, M. M. *Children from seed to saplings* (2nd ed.). New York: McGraw-Hill, 1951.
- Rivlin, H. N. *Educating for adjustment*. New York: Appleton-Century, 1936.
- Robison, R. K. Newer trends in psychoanalysis. *Progressive Education*, 1949, **27**, 49-51.
- Robison, R. K. Neo-Freudian psychoanalysis and progressive education. *Progressive Education*, 1954, **31**, 89-90.
- Ryan, W. C. *Mental health through education*. New York: Commonwealth Fund, 1938.
- Ryan, W. C., & Skinner, C. E. (Eds.). *Elementary educational psychology*. New York: Prentice-Hall, 1945.
- Sandiford, P. *Educational psychology*. London: Longmans & Green, 1929.
- Sandiford, P. *Foundations of educational psychology*. New York: Longmans & Green, 1938.
- Schilder, P. Psychoanalysis and conditioned reflexes. *Psychoanalytic Review*, 1937, **24**, 1-17.
- Schrupp, M. H., & Gjerde, C. M. Teacher growth in attitudes toward behavior problems of children. *Journal of Educational Psychology*, 1953, **44**, 203-214.
- Searl, M. N. Some contrasted aspects of psycho-analysis and education. *British Journal of Educational Psychology*, 1932, **2**, 276-296.
- Sears, R. R. *Your aunts revisited: A history of child development*. Chicago: University of Chicago Press, 1975.
- Shaffer, L. F. *The psychology of adjustment*. Boston: Houghton Mifflin, 1936.
- Shaffer, L. F., & Shoben, E. J., Jr. *The psychology of adjustment* (2nd ed.). Boston: Houghton Mifflin, 1956.
- Shakow, D., & Rapaport, D. *The influence of Freud on American psychology*. New York: International Universities Press, 1964.
- Sherman, M. *Mental hygiene and education*. London: Longmans, Green, 1934.
- Simpson, R. G. *Fundamentals of educational psychology*. Chicago: Lippincott, 1949.
- Soule, G. Psychology and revolution. *New Republic*, 1937, **92**, 66-72.
- Sparks, J. N. Teachers' attitudes toward the behavior-problems of children. *Journal of Educational Psychology*, 1952, **43**, 284-291.
- Sperling, M. Psychoanalytic aspects of discipline. *Nervous Child*, 1951, **9**, 174-186.
- Spock, B. *The commonsense book of baby and child care*. New York: Duell, Sloan & Pearce, 1946.

- Starch, D. *Educational psychology*. New York: Macmillan, 1919.
- Starch, D. *Educational psychology* (Rev. ed.). New York: Macmillan, 1927.
- Starch, D., Stanton, H. M., & Koerth, W. *Psychology in education*. New York: Appleton-Century, 1941.
- Steere, G. H. Freudianism and child rearing in the twenties. *American Quarterly*, 1968, **20**, 759-767.
- Stendler, C. B. Sixty years of child training practices. *Journal of Pediatrics*, 1950, **36**, 122-134.
- Stouffer, G. A. W., Jr. Behavior problems of children as viewed by teachers and mental hygienists: A study of present attitudes as compared with those reported by E. K. Wickman. *Mental Hygiene*, 1952, **36**, 271-285.
- Strang, R. *An introduction to child study*. New York: Macmillan, 1930.
- Strang, R. *An introduction to child study* (Rev. ed.). New York: Macmillan, 1938.
- Strong, E. K. *Introductory psychology for teachers*. Baltimore: Warwick & York, 1920.
- Strong, E. K. *Introductory psychology for teachers* (Rev. ed.). Baltimore: Warwick & York, 1922.
- Stroud, J. B. *Educational psychology*. New York: Macmillan, 1935.
- Stroud, J. B. *Psychology in education*. New York: Longmans & Green, 1946.
- Sullivan, M. *Our times: The United States 1900-1925* (Vol. 4). New York: Scribner's, 1932.
- Sulman, A. M. The humanization of the American child: Benjamin Spock as a popularizer of psychoanalytic thought. *Journal of the History of the Behavioral Science*, 1973, **9**, 258-265.
- Symonds, P. M. *Mental hygiene of the school child*. New York: Macmillan, 1934.
- Symonds, P. M. The value of courses in mental hygiene for the personality adjustment of prospective teachers. *Mental Hygiene*, 1941, **25**, 568-575.
- Symonds, P. M. Education and psychotherapy. *Journal of Educational Psychology*, 1949, **40**, 1-32.
- Taft, J. Rank's contribution to education. *Progressive Education*, 1932, **9**, 495-498.
- Tanner, A. E. *The child: His thinking, feeling, and doing* (Rev. ed.). Chicago: Rand-McNally, 1915.
- Thomas, W. I., & Thomas, D. S. *The child in America: Behavior problems and programs*. New York: Knopf, 1928.
- Thompson, G. G. *Child psychology: Growth trends in psychological adjustment*. Boston: Houghton Mifflin, 1952.
- Thorndike, E. L. *Educational psychology*. New York: Lemcke & Buechner, 1903.
- Thorndike, E. L. *Educational psychology* (3 vols.). New York: Columbia University, Teachers College, 1913-1914.
- Thorndike, E. L. *Educational psychology: Briefer course*. New York: Columbia University, Teachers College, 1914.
- Thorndike, E. L. *The psychology of arithmetic*. New York: Macmillan, 1922.
- Thorndike, E. L. *The psychology of algebra*. New York: Macmillan, 1923.
- Thorndike, E. L. *Human nature and the social order*. New York: Macmillan, 1940.
- Thorpe, L. P. *The psychology of mental health*. New York: Ronald Press, 1950.
- Thurstone, L. L. Contributions of Freudism to psychology. I. Influence of Freudism on theoretical psychology. *Psychological Review*, 1924, **31**, 175-183.
- Tiegs, E. W., & Katz, B. *Mental hygiene in education*. New York: Ronald Press, 1941.
- Torrs, H. L. *Psychological foundations of teaching*. New York: Globe, 1927.
- Totem and taboo. *New Republic*, July 20, 1918, pp. 347-348.
- Tyler, J. M. *Growth and education*. Boston: Houghton Mifflin, 1907.

- Tyler, L. I. Psychoanalysis and curriculum theory. *School Review*, 1958, **66**, 446-460.
- Valentine, C. W. *Psychology and its bearing on education*. New York: Philosophical Library, 1951.
- Viereck, G. S. Freud: Columbus of the unconscious. *Forum*, 1925, **73**, 302-312.
- Waddell, C. W. *An introduction to child psychology*. Boston: Houghton Mifflin, 1918.
- Wallin, J. E. W. *The mental health of the school child*. New Haven, Conn.: Yale University Press, 1914.
- Wallin, J. E. W. *Personality maladjustments and mental hygiene* (2nd ed.). New York: McGraw-Hill, 1949.
- Watson, G. What shall be taught in educational psychology? *Journal of Educational Psychology*, 1926, **17**, 577-599.
- Watson, G. Psychoanalysis and the future of education. *Teachers College Record*, 1957, **58**, 241-245.
- Weeks, H. F., Pickens, H. D., & Roudebush, R. I. A comparative study of recent texts in psychology, educational psychology, and principles of teaching. *Journal of Educational Psychology*, 1930, **21**, 327-340.
- Weinreb, J. Report of an experience in the application of dynamic psychiatry in education. *Mental Hygiene*, 1953, **37**, 283-293.
- White, W. A. *Mental mechanisms*. New York: Journal of Nervous and Mental Disease Publishing Company, 1911.
- White, W. A. *Mechanisms of character formation*. New York: Macmillan, 1916.
- White, W. A. *The principles of mental hygiene*. New York: Macmillan, 1917.
- White, W. A. *The mental hygiene of childhood*. Boston: Little & Brown, 1919.
- White, W. A. The origin, growth, and significance of the mental-hygiene movement. *Mental Hygiene*, 1930, **14**, 555-564.
- Wickman, E. K. *Children's behavior and teachers' attitudes*. New York: Commonwealth Fund, 1928.
- Wittels, F. Psychoanalysis for teachers. *Progressive Education*, 1931, **8**, 238-241.
- Woody, C. The administration of the psychology prerequisite to courses in education. *Journal of Educational Psychology*, 1920, **11**, 61-77.
- Worcester, D. A. The wide diversities of practice in first courses in educational psychology. *Journal of Educational Psychology*, 1927, **18**, 11-17.
- Yourman, J. Children identified by their teachers as problems. *Journal of Educational Sociology*, 1932, **5**, 334-343.
- Zachry, C. B. The role of art in personality development. *Mental Hygiene*, 1933, **17**, 51-58.

Chapter 6

THE IMPACT OF THE LANGUAGE SCIENCES ON SECOND-LANGUAGE EDUCATION

by

H. H. STERN, MARJORIE BINGHAM WESCHE,
AND BIRGIT HARLEY

*Modern Language Centre
The Ontario Institute for Studies in Education*

Two commonly held views of the relationship between research and teaching are roughly as follows: First, theoreticians and researchers formulate theories, develop concepts, and carry out "pure" research within certain disciplines. The findings of these studies are "applied" in education and lead to an improvement in classroom performance. If research is funded, this will in due course pay off in better teaching and learning. Second, an educational need, demand, or deficiency is discovered by politicians, educational critics, the educators themselves, or the public. A funding body sets up a research fund that stimulates research and theory related to this need, demand, or deficiency. The findings are applied in the field of educational practice and the need is met, the demand fulfilled, or the deficiency remedied.

However, it is often found that the educational expectations associated with these two views are not fulfilled. Educational practice has not necessarily been visibly improved in spite of generous research funding and a high level of research activity. Questions then begin to be asked. Has the wrong research been funded? Have the researchers done a bad job? Should there perhaps be less emphasis on research? Has there been a defect in communication between the theoretician-researcher and the

The authors of this chapter acknowledge gratefully the generous support of Dr. F. M. Connelly, Chairman, Department of Curriculum, Ontario Institute for Studies in Education, who made available facilities for the case study. We are also grateful to C. K. Seshadri, I. I. King, and D. A. Gallagher for invaluable assistance in preparing the manuscript.

practitioner? Have the implications and limitations of research results been fully understood? Is the conception of the relationships between research and practice at fault? Is the practitioner at fault by failing to adopt what research can offer? Are there other influences at work that override the effects of research?

There is a movement in foreign-language teaching theory—and in curriculum studies generally—that rejects what it believes is the heavy hand of research disciplines. A recent book on language teaching, by Jakobovits and Gordon (1974), seeks to liberate the language teacher from the "tyranny of irrelevant expertise" claiming that "the 'expert' and his research have been elevated to totally unrealistic levels of respect and adoration" (p. 85) and that "this deference to an all-powerful research divinity is entirely misplaced" (p. 86). Jakobovits and Gordon maintain that the teacher should be able to rely on his own resources and not be dependent on the directions of experts and supervisors:

A clear distinction is to be made between "basic research" and "applied research." Basic research is esoteric, specialized, and inaccessible to the teacher. To be a consumer of it, to be dependent upon it means to be subjugated to the authority and expertise of others. It means giving up the freedom of choice in favor of faith and trust in the technocrat who very often is far removed from the realities and needs of the classroom, and in any event, is not the person who is held "accountable." (p. 93)

Having liberated the teacher from this external tyranny, Jakobovits and Gordon do not completely reject the techniques of the research discipline but advocate that the teacher should apply these himself in search of his own solutions. That is to say, the teacher as a fully responsible professional person applies whatever teaching or research technique is appropriate in a given situation (e.g., p. 96). The practitioner is no longer viewed as the passive recipient of an external expertise. He himself becomes the "expert" who makes use of basic knowledge, theories, concepts, or research findings as needed in policymaking, teacher training, materials development, and so on.

This approach is similar to the rejection of the "theoretic" and its replacement by deliberation (or planning) as a preparation for policy and decision making, which has been recommended by Schwab (1970) as the appropriate "language" for curriculum. But, as Roberts and Connelly (1975) have pointed out, in order to deliberate, the teacher must have something to deliberate about, and he must have the necessary conceptual and analytical tools with which to work.

However, the practitioner frequently lacks the theoretical background and access to the information he needs to use these resources adequately. The education of the language teacher, therefore, and the intermediaries between fundamental research, theories, and language pedagogy assume special importance in developing this relationship and are ultimately vital for the improvement of language learning.

This case study illustrates the conditions that need to obtain for research and an educational activity to be mutually beneficial.

There are a number of reasons why second- or foreign-language teaching is a good choice for a case study of this issue: (a) Language learning has baffled educators for centuries; (b) there has been a research tradition for over a hundred years; (c) there have been vigorous developments over the last 30 years in linguistics, psycholinguistics, and sociolinguistics linking these disciplines to language teaching practice.

The purpose of this chapter, therefore, is to explore the impact of research in the language sciences on the field of language teaching.

Research in this context includes studies in the language-related disciplines of linguistics, psycholinguistics, and sociolinguistics, as well as a few examples of studies undertaken within the framework of educational research. In addition to empirical and experimental investigations, we have taken into account theoretical developments that we or other applied linguists and language teachers have regarded as important for language teaching. We have not confined ourselves to studies prompted by the wish to improve or expand second-language¹ instruction.

Being selective, this chapter should not be read as a full appreciation of all aspects of research that might have relevance for language teaching. It is not a comprehensive research review, but rather an attempt to focus on, by way of example, some studies, aspects, and trends that in our estimation throw light on the development of a relationship between the language sciences and language teaching.²

Considering the question of how to demonstrate conclusively the *impact* of research upon second-language instruction, we found ourselves in sympathy with the point of view expressed by Clifford (1973): "To establish the facts of the impact of educational research upon teaching itself, to advance quite certain proof, and to reconstruct a relationship of cause and effect is a near impossibility. If either the praise (including contracts, royalties, promotions) or the blame accorded educational research depended on the existence of hard data on its impact upon schools, there would be little of either" (p. 3).

In considering what would constitute an impact, we came to the conclusion that the ideal pragmatic test would be to show that as a result of research in, say, some aspect of linguistics, the quality of instruction was noticeably enhanced, its availability increased, or the achievement of its students better than that of students taught without the benefit of the research input. But achievement in a language is dependent on many

¹ Throughout this chapter the phrase "second-language (or L2) teaching" is used as a general term to refer to all kinds of formal instruction in languages other than the "mother tongue." For our purposes here, it is not necessary to make a distinction between "second" and "foreign" languages. The terms "language pedagogy" and "language education" also refer to second-language instruction.

² For research reviews, see, in section 6.2, the subsection entitled *Documentary sources*.

factors. If we regard achievement in learning as dependent on the factors identified by Carroll in his model of school learning (Carroll, 1963a), and if we add to these the factors that he has identified as specific to foreign-language learning (Carroll, 1969a), it would be difficult or probably impossible to isolate the specific effect of research.³

On the other hand, we can look for the impact of research in other directions: In addition to exercising an influence on the classroom practitioner and student, research might affect the policymaker, curriculum developer, writer of teaching materials, supervisor, adviser, language-teacher trainer, or the general public.

The nature of the impact may be in the form of *information* that as a result of research appears in the design of course materials, in teaching strategies, and so on. For example, it is noticeable that in older, "prescientific" teaching grammars very little distinction was made between spoken and written language usage. To the extent that teachers or textual materials now make this distinction, linguistic research has made an impact.

Research may also show itself in a more abstract form in the *theory* of language or language learning or in the *concepts* of language pedagogy. Thus, if a teacher makes a distinction between his students' competence and their performance, this formulation is likely to be influenced by the conceptual distinction made in linguistics by Chomsky (1965).

The influence of research may also be in providing the practitioner with *explanations* that create a conscious understanding of something he previously felt only intuitively. It is not necessary for the findings to be totally new; they may simply be an aid to making sense of what is happening in language teaching (Clifford, 1973). For example, language teachers have for decades been torn between focusing their teaching on linguistic forms or on meanings and the overall context within which the language is used. Likewise, the different ways in which the research focus can be restricted to linguistic form or widened to include meaning and social context have been major issues in the development of linguistic theory and research. What has been intuitively felt by language teachers has been conceptualized by linguists. To the extent that this conceptualization is recognized as the same dilemma in language teaching, one acknowledges an impact: Linguistic theory leads to heightened awareness.

More indirectly, research can influence *attitudes* toward the problems of language teaching: It invites a questioning outlook, suspension of judgment, caution, and a search for evidence.

³ To the factors included in the model of school learning, Carroll (1969a) has added for foreign-language learning: the difficulty of the language to be learned; the level of attainment in the foreign language expected; the opportunity to learn the language in school and the opportunity for contact with the second language in school and outside of school; the social status of the language and its cultural, political, or instrumental value; and the mode of learning the language as a formal system or through communication.

Lastly, research can have a bearing on *decision making* in language teaching. Under this heading, research rarely speaks with absolute authority: It puts options before the decision maker. It may indicate how much is known and to what extent knowledge is assured. One of the indictments made of existing language-teaching research is that on crucial issues—for example, on the question of an optimal age for language learning, or on methods of language teaching—it has not given the decision makers sufficiently firm knowledge.

In this study, we begin with a brief historical sketch of the relation of research and language teaching, in order to give some perspective to current developments, and follow with a short description of four possible models of interaction. Next, recent linguistic and sociolinguistic influences are examined, followed by a consideration of psychological and psycholinguistic research. A short section concentrates on a few examples of educational research. We also give attention to organizational mechanisms through which the language sciences and language teaching have been related to each other. In the final section we attempt to evaluate the interaction between language research and language teaching.

The chapter is based mainly on developments in North America, but also considers broad international trends. The relationships between the language sciences and language pedagogy in Europe and the Soviet Union have been discussed in detail by Lewis (1974), while several reports by the Ford Foundation (e.g., Fox, 1974, 1975; Ford Foundation, 1975) provide information on other parts of the world.

1. HISTORICAL ORIENTATION

Language teaching has a tradition reaching back for 25 centuries (Kelly, 1969). Early in its history, it developed its own techniques based on the intuitive interpretation by its teachers of what was needed to teach a living language. Stemming from the development of formal education in the 18th and 19th centuries, a more self-conscious language pedagogy has evolved, particularly since the 1880s. At about that time, a vigorous reform movement began in Europe aiming at radical improvement of language-teaching processes. This movement coincided with the early development of two disciplines: one studying language, namely, linguistics; and the other studying language learning, namely, psychology. Over the last hundred years, repeated attempts have been made to give language teaching a basis in the theory and research derived from these two disciplines. Thus, language education, like education in general, has been increasingly seen as an area in which scientific research can be usefully applied.

1.1 1880 to 1920

The first example of a direct link between linguistics and language teaching is the influence of phonetics from about 1880. Phonetics was seen as a means of bringing the spoken language and authentic descriptions of the pronunciation of the foreign language into the classroom. The development of the International Phonetic Alphabet (IPA) and its introduction into the language class were considered by some a real breakthrough.

In subsequent decades, the enthusiasm for phonetics and IPA died down. Yet, phonetics remains one of the most powerful influences of linguistic research on language teaching. Gimson's *An Introduction to the Pronunciation of English* (1970), Prator's *A Manual of American English Pronunciation* (1957), and work by Leon (1966) on the pronunciation of French are direct descendants of the earlier studies on phonetics. Such books are reference works for the course writer, teacher trainer, or teacher who wants to understand what he is doing, rather than texts for direct use in the classroom. The lasting influence of phonetics has thus been indirect, while the direct influence—the teaching of the phonetic alphabet and use of diagrams of the vocal organs—has become less prominent or disappeared. This is not surprising. A phonetic notation or a vowel chart may be an excellent research tool but is not necessarily a good pedagogical device.

With historical hindsight, an interesting and puzzling observation is that phonetics alone was regarded as the scientific input from linguistics. Today this seems incomprehensible. Why no other aspect of linguistics? Why not grammar or vocabulary? In practice, language teaching was very much concerned with grammar and vocabulary. But these were treated simply as subjects for commonsense analysis. Sensitive language-teaching theorists like Gouin (1892) and Palmer (1917/1968) developed their own comprehensive linguistic theories because they felt the need for it, but they had no substantial body of theory or research to draw on.

Similar observations can be made about the beginnings of psychology. Psychological influences are recognized as relevant from about 1880. Psychological thought at that time provided a conceptualization of intuitively felt psychological processes. The British linguist Sweet (1899/1964), for example, recognized that the "law of association" accounted for mental processes he regarded as important in language learning. This "law" was also one of the main psychological justifications for the "direct method" of language teaching advocated at that time. Sweet developed his principles of association on the basis of his intuitions and teaching experience; they were never empirically verified with real students. Here we meet for the first time a problem that the application of psychology to language teaching has brought with it: oversimplification, a phenomenon

that has presented itself again and again in the attempt to use scientific knowledge in solving language-teaching problems.

1.2 1920 to 1940

From the 1920s to the 1940s, language teaching was interpreted mainly in terms of educational psychology, which at that time was developing vigorously. There was no direct study of the second-language learning process itself; it was principally a matter of applying to second-language teaching the psychological concepts and research techniques developed elsewhere. The possibility of a psychological interpretation of learning was recognized, but little was done to map out where existing psychological theory could make a specific contribution. Language teachers talked about characteristics of language learners and the conditions of learning on the basis of practical teaching experience alone, without recognizing the link between such experience and a possible psychology of second-language learning.

Psychologists, for their part, hardly recognized that second-language learning might be a psychologically interesting phenomenon, although first-language acquisition and bilingualism were studied as a part of child development. Equally, few theoretical linguists were aware of the potential linguistic interest of problems that were arising in second-language teaching.

An impressive amount of research was nonetheless undertaken between 1924 and 1930 by the American and Canadian committees on foreign-language teaching. These studies mark the beginning of empirical research in language education, based largely on research methods used in educational psychology. In particular, the application of psychometrics led to the development of objective language achievement tests and attempts at a language prognosis test around 1930 (e.g., Symonds, 1930). Characteristically, this early test development was viewed exclusively as educational research or as research in educational psychology. That test development was also a linguistic task, which implies a theory of language, was not yet evident.

Language testing was taken up again at the English Language Institute at Ann Arbor and by the Educational Testing Service, but it remained essentially an application of educational testing until the mid-1950s.

Another example of how "nonlinguistically" language was conceptualized by educators in the interwar years is the word- and syntax-frequency studies initiated by the Modern Foreign Language Study in the 1920s. The pattern was set by the eminent educational psychologist of that period, E. L. Thorndike, who with characteristic directness and simplicity produced *The Teacher's Word Book* in 1921.

In sum, language in education—be it the mother tongue or a foreign language—was treated, during this period, as a phenomenon that could

be appropriately and adequately investigated by educational psychologists, but not as something to which a developing linguistic science had anything to contribute.

1.3 1940 to 1960

Since 1940 there has been a marked change. The American linguistic tradition known as structuralism—led in its early stages by Bloomfield—was, through force of circumstance, to have a major impact on language education. Starting in the 1920s, anthropological linguists working on the description of American Indian languages and other unwritten tongues developed elaborate methods to record these spoken languages in the field and to analyze them in terms of their formal aspects: phonology, morphology, and syntax. Through their field investigations, they began to see the relevance of their methods and descriptions for language pedagogy (Bloomfield, 1933).

With the involvement of the United States in World War II, it became apparent that large numbers of Americans would need practical conversational skills in languages that, in many cases, were not taught in U.S. schools and universities. The existing language departments in the universities, oriented toward the humanities and literature, were not equipped to respond to the new situation. Through a series of circumstances, linguists soon became heavily involved—not only in providing descriptive materials for the needed languages, but also in the practical organization of language training courses on a massive scale. This new role for linguistics in language teaching from 1940 to 1960 is described in detail by Moulton (1961).

In 1941, through an initiative of the American Council of Learned Societies (ACLS), an "Intensive Language Program" was set up in conjunction with an ACLS affiliate, the Linguistic Society of America. Funded by the Rockefeller Foundation, linguists in the program prepared analyses of a number of languages for which no teaching materials existed using Bloomfield's *Outline Guide for the Practical Study of Foreign Languages* (1942), written especially for the program, and Bloch and Trager's *Outline of Linguistic Analysis* (1942). The armed forces turned to this program and, by the end of 1943, some 15,000 soldiers were studying 27 different languages at 55 universities (Moulton, 1961). The American wartime experience in intensive language training was ultimately to involve, in one way or another, almost every trained linguist in North America. Foreign-language pedagogy and linguistics, thus thrust together, would have far-reaching consequences.

The methodological approach developed for the wartime courses, largely following Bloomfield, had three main characteristics: (a) Courses were based on structural linguistic descriptions; (b) linguistic field methods were introduced into the classroom, including the use of native

informants and genuine samples of the spoken language; and (c) teaching methods were based on a behaviorist theory of learning as interpreted by the structuralists, who viewed language as "a system of extremely well-learned habitual responses" (Carroll, 1953, p. 191; see also Bloomfield, 1933). The methodology resulting from this psychological interpretation of language learning incorporated the use of imitation, repetition with selected reinforcement, and drill and practice, with "overlearning" as the goal. This approach, which later became the "audiolingual method," came to dominate the thinking of many language educators in the 1950s and 1960s.

Thus, from the early 1940s to the present, there is no doubt of the impact of linguistics on language teaching, and, increasingly, of psychology. In return, the scope of the concerns of both disciplines widened, reflected in the development of the related fields of applied linguistics and psycholinguistics. Together, these disciplines began to deal more comprehensively and systematically with the needs and issues of language education.

After the end of World War II, the links between linguistics and language education were maintained by the Committee on the Language Program (CLP) of the ACLS through its fellowship programs, summer institutes, and publications. Some of the language texts used in wartime programs were commercially available to the public, notably the "Spoken Language" manuals, published by Henry Holt and Company. Features of the "Army Method" were retained in postwar language courses at a number of universities, particularly for the "new" languages. A notable example was Cornell University, which reformed its entire modern-language program to include beginning courses stressing conversational skills based on the army method and offered linguistics courses for foreign-language majors. In some universities, however, the traditional language departments remained largely untouched.

The new methods and goals of language training were also prominent in the Language Training Program of the Foreign Service Institute, begun in 1946, and at the Army Language School in Monterey, California. Both institutions quickly became active in curriculum development and research in language education and continue to contribute to the field today.

At first, secondary-school language programs were not much affected by these developments. Many teachers remained convinced that the new methodology worked only under intensive training conditions, where classes were very small and where the primary goal of study was an oral command of everyday language. But gradually, during the 1950s, through increased familiarity with the new methodology, experience in the universities, and new curriculum materials, the influence began to filter down (see discussion in Moulton, 1961).

What linguists felt to be the prime innovation in the method—the use of a scientific descriptive analysis of the target language as the basis for teaching materials (Fries, 1949)—was often not recognized by language teachers. Fries, however, who directed the English as a second language (ESL) program offered to foreign students at the English Language Institute of the University of Michigan, begun in 1941, undertook an analysis of English sentence structure for use in curriculum materials (1952). His analysis was based on recorded conversations of some 250,000 words. In addition to stressing the scientific description of authentic L2 speech as a basis for language-teaching content, Fries (1945) saw as essential “an adequate descriptive analysis of . . . the native language of the student” (p. 5). Pioneering work in “contrastive analysis,” as it came to be known, was carried on at the Institute. One of its first applications was the elaboration of separate ESL courses for different groups, including Spanish-speaking Latin Americans and speakers of Chinese.

Carroll has been the foremost psychologist to concern himself regularly and specifically with psychological questions of second-language learning. As early as 1953, in a survey of linguistics and related disciplines (Carroll, 1953), he recommended the use of carefully controlled, interdisciplinary experimental studies to provide answers to the perennial questions of language teaching, and he has continued to stress the importance of a research approach (Carroll, 1960, 1963b, 1967, 1969a, 1975b).

In 1953, a meeting of historical importance took place between psychologists and linguists at Indiana University, which contributed directly to the development of the new interdisciplinary field of psycholinguistics (Osgood & Sebeok, 1954). The widely distributed report of this meeting proposed a framework defining the role of the new discipline and its relation to psychology and linguistics. Although second-language acquisition and bilingualism received little attention at the conference itself, they were included in the framework as topics of interest. Of major significance for language pedagogy, this event represented the first serious attempt to relate second-language learning and bilingualism to a theory of language acquisition and use.

Several other events of the 1950s were important for the development of a relationship between psychology, linguistics, and language education. One of these was the publication of Skinner's *Verbal Behavior* (1957), which furthered a behavioristic theory of L2 learning. Another was Chomsky's (1959) review of this book, in which he attempted to demolish not only Skinner's thesis but behaviorism as a viable theory of language in general. However, the impact of this attack was not felt in language pedagogy for several years. A third event was passage of the National Defense Education Act in 1958, which was to provide unexpected funds

for language research and pedagogy. Yet another notable event was the creation of the Center for Applied Linguistics in Washington, D.C., in 1959.

Around 1960, there was a strong awareness, particularly in the United States, that language teaching was undergoing a revolution. One can speak of the "peak" of audiolingualism in the 1960s in the United States, while in Canada, Australia, and Europe its impact was felt a few years later. Perhaps language teachers in these countries had a longer tradition of exposure to different linguistic schools and resisted the ideological enthusiasm of some of their American counterparts toward the applications of structural linguistics that were featured in the new approach (Halliday, McIntosh, & Strevens, 1964; Mackey, 1965). Nonetheless, by the 1960s, it was generally recognized that language teaching had to have an adequate basis in both linguistics and psychology.

1.4 Trends: 1960 to 1975

During the past ten to fifteen years, developments in research and theory related to language teaching may be described in terms of three major trends. The first is the continuing influence of general and descriptive linguistics, the revolutionary impact of transformational generative grammar, and the growing recognition of the significance of new work in semantics and sociolinguistics.

The second major trend involves the recognition of the fundamental relevance of psychology and psycholinguistics for language teaching, with a growing emphasis on cognitive psychology and research on both first- and second-language acquisition.

A third trend involves policy-oriented research on language education in a school setting. Such research often brings together language educators, psychologists, and linguists in a team effort to answer major policy questions or to evaluate experimental programs.

The trends sketched here will be discussed in the following sections of this chapter. It should be noted that these trends can only be roughly divided because, in practice, they have been so intermingled that a rigid analytical separation is not possible.

2. MODELS OF INTERACTION

The history of the relations between the language sciences and language pedagogy examined in this case study reveals the problems of relating research to teaching and also shows what attempts have been made to resolve these problems. The various lines of thought that have been pursued can be summarized as four different models.

2.1 *The Separation or "Hands Off" Model*

Language teaching is one thing. Language teachers follow a tradition that has its own psychological and linguistic positions; decisions are made without reference to linguistics or psychology. Linguists and psychologists, equally, follow their own paths and do not expect their theories and researches to be relevant to language pedagogy (e.g., Chomsky, 1966).

This model has the merit of firmly recognizing that researchers and teachers fulfill different functions and theirs should not be a relationship of dependence. On the other hand, it overlooks the fact that language teaching presents problems that can be of theoretical interest to research and that linguistics or psycholinguistics may produce findings of interest to language teaching. It also fails to recognize that researchers and teachers live in the same intellectual climate. If they speak a common "language" they can share and exchange thoughts and problems; for language research and language teaching have, after all, much in common.

2.2 *The Application Model*

Linguistics and psychology provide scientific concepts, theories, and attested information that can, and should, be applied in language teaching. On the language teacher's side, there is an expectation that these concepts, theories, and data will be provided and will modify and improve language teaching. In this model, which is the most widely adopted one, the question is how best to regulate the form of application. This view was recently made explicit in Corder's (1973) *Introducing Applied Linguistics*. According to Corder the relationship is a strictly defined, one-way street: "The application of linguistic knowledge to some object—or applied linguistics, as its name implies—is an activity. It is not a theoretical study. It makes use of the findings of theoretical studies. The applied linguist is a consumer, or user, not a producer of theories" (p. 10). In this model, applied linguistics is a specific mediating activity that, in specific ways, contributes to the total language-teaching operation on three different levels. At the highest and most general level, government authorities decide whether to teach languages at all, which languages to choose, and to whom these languages are to be taught. At this level the contribution of applied linguistics is confined principally to the kind of information that is provided by language planning. Applied linguistics comes into its own at the second level that is concerned with "what to teach, when to teach, and how much to teach"; for according to Corder, applied linguistics "has to do with the devising of syllabuses and materials for carrying out the intention of education authorities whether local or national" (p. 3). The third level is primarily that of the classroom teacher where the scientific input is likely to be mainly psychological and pedagogic. In other words,

the contribution of research is strictly defined: It is sociolinguistic at the highest level, linguistic and sociolinguistic at the second level, and psychological and pedagogical at the third level.

2.3 *The Resource Model*

Linguistics and psycholinguistics systematically investigate aspects of language, language use, and language acquisition that can be regarded as a resource base for the teacher to draw on. The language teacher, in this model, expects relevance but takes the initiative in seeking out and selecting aspects of linguistics and psychology. This model is the one recommended by Allen (1974) when he writes: "As teachers we should be prepared to study linguistic grammars of all types, and to select from them whatever seems to contribute, from a teaching point of view, to what Palmer calls a 'complete and homogeneous' system of presentation" (p. 91). On the positive side, this model assigns to the teacher the active role of selecting the theories, concepts, and information he requires from what is offered by the language sciences. On the other hand, it presupposes a discriminating understanding of the whole spectrum of the language sciences, probably too demanding a supposition. In order to develop his own applications, the teacher must know what the right questions are to ask, where the information can be found, and which information is applicable and which is not. Recent history has shown that this can lead to a random picking of tidbits and to misleading misinterpretations of scientific concepts out of context.

2.4 *The Common Ground or Convergent Model*

This model combines features of the three previous ones and is the one we favor. Linguistics and psychology are recognized as activities in their own right following the impetus of their own theories and research traditions. Language teaching also exists in its own right and has its own traditions. But the language sciences have something in common with language teaching: they are all concerned with language, language acquisition, and language use. The language sciences provide theories, develop concepts, and collect and order information. Language teaching as a practical activity, for its part, has its own theories, concepts, and data. But it requires information and it reveals problems that can stimulate research and theory. Research, theory, and practical activity can support each other. The relationship is most productive if it is viewed as one of reciprocity, and if it leads to interchange between theoreticians, researchers, and practitioners on equal terms.

Bearing these different models of interaction in mind, we can now place language pedagogy in the wider context of research and its relation to education in general.

3. LINGUISTICS AND SOCIOLINGUISTICS

Aspects of research in linguistics relevant to language teaching will be discussed in the light of four roughly chronological developments: (a) the confident application of linguistics, (b) the disorienting impact of new theory, (c) a reassessment of the relationship, and (d) new directions.

3.1 *The Confident Application of Linguistics*

During the early 1960s, language teaching was characterized by a sense of progress and achievement attributed to the introduction of scientific principles to language teaching. In 1964, a book addressed to language teachers was significantly titled *Language Teaching: A Scientific Approach* (Lado, 1964). It refers to a "great movement" in language teaching, in which "dramatic advances in linguistic science" (p. 3) are seen as a major factor. In the same year, another writer (Stack, 1964) makes the claim that

today's foreign language teaching is achieving success unknown under the traditional methods. This has been accomplished by the application of structural linguistics to teaching, particularly in the realm of proper sequence, oral grammar, inductive grammar, and the use of pattern drills to give intensive practice. (pp. 80-81)

Some of the aspects of structural linguistics that were so confidently "applied" to second-language teaching are now examined. The history of their use illustrates both the merits and problems of using a simple application model.

Empiricism and the primacy of speech. Structural linguists of the 1930s to 1950s adopted a rigorous empirical approach to the study of language. Theirs was a "descriptive" approach as opposed to the "prescriptive" one of traditional grammar.⁴ By objective analysis of a carefully recorded "corpus" of authentic language data gathered from a native speaker, the structuralist aimed to arrive, by induction, at an explicit description of a language on its terms and not in the mold of traditional grammar based on the categories of classical Greek and Latin.

Since the structuralists were often concerned with unwritten languages, they were led, in contrast to traditional grammarians, to emphasize the primacy of speech over writing. Speech is the primary means of communication in all societies, they pointed out, while writing is less widespread and is clearly based on spoken elements of language. According to Bloomfield (1933), for example: "Writing is not language, but merely a way of recording language by means of visible marks" (p. 21). In addition,

⁴ It was the speech of the native-speaker informant and not some remote literary standard of correctness that determined what was an acceptable utterance in the language under investigation.

structuralists concluded that since children learn to speak their mother tongue before they learn to write it, speech before writing represents the natural order of language acquisition.

The linguistic emphasis on speech is reflected in many language-teaching materials that were produced during the early 1960s and that are still widely used today. Dialogues, that is, simulated spoken language, largely replace written narrative prose. Often, the introduction of reading and writing is delayed for a considerable period of time, and the acquisition of reading and writing skills is viewed essentially as a matter of associating written symbols with spoken sounds. However, it would be an oversimplification to say that the emphasis on spoken language derived solely from linguistics. It may also be seen as a pragmatic response to the newly felt language-teaching goal of communication between peoples (e.g., Jones, 1964, pp. 40-41) and the evident failure of traditional texts—with their emphasis on writing, translation, and reading of foreign literature—to produce students able to speak the foreign language. While linguistic theory in Britain and Europe did not derive from Bloomfieldian structuralism, opposition to traditional grammar, along with a new emphasis on spoken language and the empirical description of languages, was also evident there (e.g., Halliday et al., 1964).

The development of *Le Français Fondamental* (1^{er} degré)⁵ during the 1950s is a European example of how such linguistic notions were brought to bear on a major research project, the findings of which have been extensively used in French L2 teaching materials. The aim of this study was to produce a list, as a basis for beginning-French materials, of the most frequently and widely used words and structures in current spoken French. Initial data consisted of tape-recorded conversations on everyday topics with 275 native speakers of French. The "informants" represented different age groups from a range of regional, social, and educational backgrounds. A further study, undertaken to determine the most readily "available" words on a number of familiar topics, involved asking groups of school children to write down the 20 words that first came to mind on a given topic. A final list of 1,475 words was then drawn up according to the criteria of frequency, availability, and "rational empiricism." A list of structures was also drawn up by somewhat less explicit means. A number of French beginners' courses, such as *Voix et Images de France* (CREDIF, 1962), were explicitly based on *Le Français Fondamental*, although not always adhering to the original word and structure list.

Concentration on "form" and formal analysis. In keeping with their empirical approach to language, the structuralists deliberately limited their efforts to the observable plane of "form" in language, as opposed to that of "meaning." While fully recognizing the importance of meaning for

⁵ Described in Gougenheim, Michéa, Rivenc, and Sauvageat (1964).

communication, they did not regard it as susceptible to scientific investigation (Bloomfield, 1933, p. 139).

Each language was seen as having its own system of sounds (phonology) and grammar (morphology and syntax). A major concern of the structuralists was to refine the tools of linguistic analysis, whereby the units of the formal system, the phonemes and morphemes, could be objectively "discovered." The procedure involved isolating segments and substituting segments in the same linguistic environment or "frame," and observing whether any contrast in meaning was thereby effected. For example, the change from *pit* to *bit* in English involves a change in meaning; the segments /p/ and /b/ can therefore be established as phonemes. *Pit* and *bit* are referred to as a "minimal pair," differing in only one phoneme. The use of "minimal pairs" to drill phonological contrasts in the language is a common feature of audiolingual texts.

In keeping with the structuralist division of language into separate levels, lesson material was often divided into phonological and grammatical exercises. The psychological view of language as a set of habits was also applied in the extensive use of "mimicry and memorization" of dialogues in the L2. In addition, as Wilkins (1972) has noted, the Bloomfieldian preoccupation with form and the tendency to neglect semantic relations and the social function of language provided a view of language that has dominated language teaching in recent years. While oral pattern practice is not a new idea in language teaching,⁶ its wholesale adoption and systematic use during the 1960s as the major means of teaching grammatical features was new, reflecting partly the influence of behaviorist learning theory and partly the preoccupation of linguistics with form and with methods of formal analysis. The patterns are frequently designed to be drilled to the point of automaticity. In one type of drill the students are taught to manipulate formal contrasts in the language orally by substituting key elements in the appropriate structural slots in sentence frames. Each new element (or group of elements) is presented as a stimulus, either orally by the teacher, or in written form, or on tape in materials designed for language-laboratory use. The aim is to have the students learn the grammatical patterns of the language by "induction" from the language data presented. If a "rule of grammar" is ever given, and often it is not, it is in the form of a summary statement after the pattern has been mastered through repetition and drill. Such drills are a characteristic feature of second-language texts produced during the 1960s. While the overemphasis on pattern drills has rightly been criticized in recent years, such drills can nonetheless be regarded as an important step forward in

⁶ Palmer (1917/1968) had suggested drills and "substitution tables" as early as 1917. It is interesting to note that Palmer also foreshadowed structuralist views on, for example, the primacy of speech and the view of language learning as a habit-forming process (Palmer, 1922/1964).

language teaching because of their systematic and flexible approach to relevant language features. Moreover, formal structural analysis provided language teaching with an entirely new, simple, down-to-earth way of handling the complexities of a language system.

Contrastive analysis. Intuitively applied as a practical necessity in language pedagogy for many decades, a contrastive analysis of the target language and the students' mother tongue was elevated, by the introduction of precise structuralist techniques, to a systematic comparison of two languages. While structuralism provided the analytic tools for comparison, behaviorist theory provided an explanation for the generally recognized problem of "interference" from the mother tongue. The contrastive principle seemed to be a straightforward empirical answer to a host of language-teaching problems: the difficulty of new languages for learners with a given language background, interference, and grading and sequencing in the language curriculum.

Around 1960, contrastive analysis was recognized as a cornerstone in a scientifically based language pedagogy. The newly founded Center for Applied Linguistics commissioned a series of contrastive studies of English with the five commonly taught foreign languages: French, German, Italian, Russian, and Spanish. A number of second-language texts based on the contrastive principle also appeared. An example is *Modern Spanish* (Bolinger, Ciruti, & Montero, 1960), a college textbook in beginning Spanish produced by a committee of the Modern Language Association (MLA) of America. Differences between the two languages, seen as predicting points of learning difficulty for the students, are emphasized, explained, and drilled. Another text (Léon, 1964) contains exercises designed to help students from different linguistic backgrounds with specific problems of French pronunciation. Articulatory charts and phonetic descriptions are supplied along with exercises that practice minimal phonological contrasts such as *Lit/Lu/Loup*. While this particular set of contrasts in French is necessarily practiced by all linguistic groups, special remarks and exercises are reserved for individual groups. For example: "Les Scandinaves et les Germaniques, bien que possédant un 'u' dans leur langue devront faire attention à bien arrondir les lèvres pour le 'u' français, qui est plus labial que le leur" (p. 15).

Application to testing. Structural and contrastive linguistics, together with psychometrics, had a significant impact on language tests, as is illustrated by Lado's (1961) influential book on language testing. Lado was among the first to suggest that the content of language tests be based on a linguistic analysis in structural terms. While he recognized the need to test language performance as a set of integrated skills, he recommended that the test maker isolate the linguistic features (the units of sound and grammar) of the language in question and test them one by one. Standardized second-language tests produced during the 1960s embody these

principles of test construction.

Rationale for teaching methodology. In addition to its impact on teaching materials and testing instruments, structural linguistics had a considerable influence on teachers via teacher-training texts (e.g., Brooks, 1960; Politzer, 1960; Lado, 1964). From a set of principles such as "speech before writing," "authentic language standards," and "patterns as habits," Lado, for example, proceeds to outline a complete set of instructions for a "linguistic" or audiolingual method of language teaching. In commenting on Lado's book, Wardhaugh (1974) notes that it is "a deliberate attempt to formalize in extremely simple terms the prevailing views of linguistics and of psychology, and to integrate these into a statement about pedagogy." However, he cautions that "it could equally well be argued that Lado's statement about language teaching is a rationalization or justification of a set of practices that had grown up unsystematically and accidentally rather than a rigorous statement of axioms and derivative practices" (p. 132). Wardhaugh thus points to the frequent difficulty of *proving* the exact source of practices in language teaching.⁷

The impact of the "linguistic" approach on language teaching was nonetheless felt to be so pervasive that, in 1971, Norris, reviewing the field of teaching English to speakers of other languages (TESOL), was led to comment: "Contemporary approaches to language teaching are so closely attached to linguistic science that it often seems as though language teaching is a sub-division of linguistics" (p. 11).

3.2 *The Disorienting Impact of New Theory*

Even as principles of structural linguistics were being translated into practice in the classroom, transformational generative grammar—which was to shake the foundations of structuralism (and thus, by implication, audiolingualism)—was making its appearance. Although Chomsky's *Syntactic Structures* appeared in 1957 and his celebrated attack on behaviorism in 1959, it was not until the mid-1960s that any real impact was felt in the language-teaching field. By that time, Chomsky's original standard model of transformational grammar had already undergone some important modifications (Katz & Postal, 1964; Chomsky, 1965). Since then there has been constant change in theory as two major branches have developed: (a) those who, following Chomsky, posit a syntactic base component for their model of language, and (b) the "generative semanticists," such as McCawley (1968), who posit a semantic base component. This rapid

⁷ Wardhaugh continues by noting that even if it were possible to show that audiolingual practices were not derived from theoretical assumptions in linguistics, practices in language teaching need to be justified by recourse to theory, since teachers must "believe in what they are doing." Quinn (1974) expresses another point of view. He is critical of practitioners who find it tempting "to seek intellectual legitimacy and prestige by claiming descent from the dazzlingly brilliant world of linguistics" (p. 332).

evolution has served to emphasize the essentially tentative and hypothetical nature of linguistic theory (Selinker, 1972/1974; Wardhaugh, 1974) and, coupled with its abstractness and its formalism, it has discouraged those in the field of language teaching from too-ready applications. Indeed, Chomsky himself (1966, p. 43), in view of the theoretical "flux and agitation," recommended caution on the part of teachers and expressed skepticism about the significance for language teaching of "such insights and understanding as have been attained in linguistics and psychology." Here, then, we see a "hands-off" approach, or at least a demand for great caution in using an application model to relate linguistic research to language teaching.

With his attack on structuralism and his refusal to suggest applications for transformational grammar, Chomsky left language teachers with no firm theoretical ground to stand on. The sense of disorientation that ensued certainly dampened the enthusiasm of those who would have a "linguistic" method of language teaching, but it was beneficial in promoting a serious rethinking of the issues and of the relationship between linguistics and language teaching.

The creative aspect of language. One major criticism of structuralism has been aimed at its behaviorist orientation. "Linguists have had their share in perpetuating the myth that linguistic behavior is 'habitual' and that a fixed stock of 'patterns' is acquired through practice and used as the basis for 'analogy'" (Chomsky, 1966, p. 44). In Chomsky's view, language behavior is not habitual but characterized by "rule-governed creativity" involving "innovation, formation of new sentences and new patterns in accordance with rules of great abstractness and intricacy" (p. 44). The native speaker of a language has internalized a "generative grammar," which is defined as a finite system of rules that generates an infinite number of grammatical sentences and no ungrammatical ones. The linguist's task is to discover the rules of the generative grammar and its underlying principles of organization. The discovery procedures by which he arrives at such rules are impossible to specify rigorously. "One may arrive at a grammar by intuition, guess-work, all sorts of partial methodological hints, reliance on past experience, etc." (Chomsky, 1957, p. 56). Thus transformationalists call into question both the psychological and the linguistic principles that have provided the rationale for pattern-practice drill in second-language teaching.

Likewise, creativity with language is recognized as important evidence of language learning. As Wilkins (1972), for example, points out: "If we do not know the rules of language well enough to be able to make new sentences, we have not yet learned that language" (p. 219). Therefore, it is the learner's creativity that needs to be investigated when his language ability is being tested, and not his ability to utter remembered phrases and sentences.

Deep structure and surface structure. Chomsky castigated both structuralism and traditional grammar for their treatments of syntax. While traditional grammar is concerned with the relationship of form and meaning, it has been accused of suffering from inexplicitness, unclear and inappropriate categories, confused explanations, and a concentration on exceptions rather than rules (Roulet, 1972). Structuralists, on the other hand, in their concern with the data tend to concentrate on the surface structure of language. In Bolinger's opinion (1972): "The empirical bias thus postponed indefinitely looking at language as a complete system in which sounds, morphemes, syntax and meaning are mutually dependent" (p. 112). The effect is "a concentration on segments and a neglect of much syntax and most meaning."

In contrast, transformational grammar endeavored to come to grips with language as an integrated *system* of forms and, since 1965, to make explicit the relationship between form and meaning. Basic to transformational grammar is the distinction that is made between deep and surface structure. Deep structure, for Chomsky, is an abstract level of syntax that is converted into surface structure—concrete sentences—by means of formal operations called transformations. Through the deep/surface distinction the grammar is able (a) to make explicit the difference between superficially similar sentences, (b) to account for ambiguous sentences, and (c) to show how sentences that are superficially distinct may be related. In addition, by positing separate syntactic and semantic components with their own rules, Chomsky, in his 1965 model, indicates a way of accounting for deviant sentences—those that break syntactic rules and those that break semantic rules. Although in practice this distinction has proved very difficult to define, the notion of different kinds of deviancy has relevance for the analysis of errors made by second-language learners. The deep/surface distinction has, further, provided language teachers with new insights into relationships between sentences previously unformulated.

Competence and performance. Like others before him, Chomsky (1965) makes a fundamental distinction between knowledge of a language and its use in concrete situations, that is, between "competence" and "performance." Competence is "the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance" (p. 4). The structuralist, according to Chomsky, misguidedly concentrated on the observable data of performance, the surface structure of language, which is not, in his view, a direct reflection of competence because "a record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on" (p. 4).

Chomsky's view of the degenerate nature of speech data and his corresponding lack of concern with surface structure have been linked to a

diminishing preoccupation with spoken language in L2 teaching (e.g., Bolinger, 1968) and an earlier introduction of writing. However, this trend may also be a result of teachers' classroom experiences with audio-lingual materials.

The distinction between competence and performance makes explicit a problem that has arisen in language teaching from structural drills. What is it that the students know, for example, when they are able to do a drill correctly? Sometimes it appears that they have not internalized an underlying "rule" in that they are not able to use the patterns creatively in other contexts (Rivers, 1968, p. 47). Therefore they have not acquired competence and cannot be said, in the Chomskyan sense, to be truly "performing" in the language.

The goals of linguistic theory. Chomsky (1965) believes the immediate goal of linguistic theory to be the construction of a rational (deductive) hypothesis about linguistic competence. "There seems to be little reason to question the traditional view that investigation of performance will proceed only so far as understanding of underlying competence permits" (p. 10). In constructing the theory, he relegates the empirical data of language to the role of confirming or disconfirming particular hypotheses. Thus, Chomsky's goal is very different from the empirical, descriptive one of the structuralists. It has led to a new emphasis on theory-building in linguistics that helps explain why transformationalists have not, so far, embarked on comprehensive descriptions of individual languages. And yet, from the point of view of language teaching, description is seen as precisely what is most needed from linguistics (e.g., Jarvis, 1972; Quinn, 1974). In other words, a legitimate development in linguistics has led to a neglect of something important for language pedagogy.

Universalism versus particularism. In contrast to the structuralists, Chomsky emphasizes the underlying similarity of different languages and in this sees himself as carrying on the tradition of universal grammar of the 17th and 18th centuries. He sees deep-seated formal universals as implying that all languages are cut to the same pattern, though there may not be point-to-point correspondences between particular languages. In addition, he considers that there may well be "substantive" universals at all levels of language—such as distinctive features in phonology, grammatical categories, and, at the semantic level, a system of "attainable concepts."

One of Chomsky's most controversial stands concerns the "innateness" of the linguistic universals he outlines. To account for the rapidity and uniformity with which the child learns the complexities of a mother tongue, he attributes "tacit knowledge of these universals to the child" (Chomsky, 1965, p. 27). Without such a capability, he maintains, language learning "would be impossible." Along with his view that the use of language involves a process of creative construction, this emphasis on what the language learner knows has led to a new wave of research in

psycholinguistics that concentrates on the *positive* contribution of the learner in the language-learning process.

The transformational generative emphasis on universals in language has provided a new perspective on contrastive analysis. It has led to the accusation that contrastive analysis based on structuralist principles is theoretically inconsistent. The structuralists had pointed out that each language has its own system, to be described on its own terms, yet the very notion of contrast presupposes underlying universals in linguistic categories and meanings that enable comparisons. In the mid-1960s, existing contrastive analyses of the surface structures of different languages were criticized by transformationalists as superficial and theoretically untenable. At the same time, other problems contributed to a general disaffection with contrastive analysis. In spite of some remarkable *tours de force* (e.g., Moulton, 1962; Stockwell & Bowen, 1967), the contrastive description of different languages proved a much more formidable task than had at first been envisaged. Also, the new field of error analysis (see Sec. 4.3) was suggesting that contrastive analysis could not predict language-learning difficulties as fully as had been claimed. More recently, there has been a renewed interest in contrastive analysis. Reconsidered in the light of transformational theory, it has been found both theoretically defensible and practically useful (e.g., Di Pietro, 1971; Nickel, 1971; van Buren, 1974), while the earlier claims for contrastive analysis as a cure-all in language teaching have been toned down.

Transformational grammar and L2 materials. Despite the major changes in theoretical perspective, the impact of transformational grammar on L2 teaching materials has so far been less substantial than might appear. For example, a text may claim to use "transformation" drills described as syntactic operations derived from transformational grammar: addition, deletion, substitution, and permutation. But such operations were already a feature of the early audiolingual texts and even earlier direct method texts; and the conversion of written sentences from active to passive, affirmative to negative, etc., was a familiar aspect of traditional texts. Moreover, "transformation" drills are from one sentence to another, while according to transformational grammar, transformations operate on abstract underlying strings, converting them from deep to surface structures.

Equally, in second-language texts, the term "generate" has been used about a speaker to indicate that he "produces" sentences (e.g., Rand, 1969). However, Chomsky (1965) uses the term in the mathematical sense of "enumerate" or "state explicitly." It is the grammar (not the speaker) that is generative in the sense that it "explicitly states" a structural description for any grammatical sentence in the language. In the face of continuing misinterpretation, he has emphasized that

a generative grammar is not a model for a speaker or hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language by a speaker-hearer. . . . When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient way, to construct a derivation. (p. 9)

Despite the fact that the psychological status of transformational models of grammar is by no means clear (Lamendella, 1969), tentative suggestions have been made (Banathy, Trager, & Waddle, 1966; Politzer, 1972) for grading and sequencing teaching materials on the basis of Chomsky's 1957 model. The distinction between deep structure and surface structure has suggested further ways of organizing materials (Politzer, 1972); for example: different sentences with common deep structures presented together; similar sentences with different deep structures kept apart. In addition, Wilkins (1972) notes that

sometimes a transformational generative description brings together syntactic information which was previously, at best, available only in widely separated places in the grammar, and the demonstration that hitherto unconnected parts of the grammatical system are in fact related is highly informative, not only in the description of that language, but also in developing an understanding of its acquisition. (p. 91)

As an example he gives Rosenbaum's (1967) analysis of predicate complement constructions in English.

A few texts for advanced students have appeared that represent a genuine concern to incorporate insights from transformational grammar. Rutherford's *Modern English* (1968) is a case in point. It is intended for use by college and university students whose native language is not English. The transformational generative orientation is outlined in the preface, but just how far this text differs from audiolingual ones is an interesting question. A number of audiolingual features still remain—the dialogues, passages for memorization, inductive grammar, and the extensive use of drills—despite the assertion that language is not “reinforced habit.” Most important of all, the emphasis is still on language as a formal system of structural contrasts, with little indication of how the “options” provided by the grammar might be appropriately used in extended discourse. Certainly, in keeping with its intended advanced audience, the text represents a more rational, intellectual approach to L2 teaching than earlier audiolingual texts. In bringing together surface structures related by a common deep structure, Rutherford introduces a vast amount of detailed information about the syntax of English; there is much more explanation, more varied and open-ended (creative) drills, and more emphasis from the beginning on reading and writing. Overall, however, the text represents an evolution from audiolingualism; it is not the radically new

departure its preface tends to suggest.

Transformational grammar and L2 teaching methodology. The impossibility of "applying" transformational grammar directly as a method of language teaching is clearly revealed in the confused methodological debate that has arisen.

In an article entitled "How Not to Interfere with Language Learning," Newmark (1966/1970) points to a limitation that transformational theory shares with structuralism: its preoccupation with the formal structure of language and its lack of concern with use. The most important thing to teach a language learner, however, Newmark says, is how to use the language and "create new utterances that are appropriate to his needs as a language user" (p. 224). Newmark further claims that "despite our ignorance as experts, as human beings we have always known how to teach other human beings a language: use it ourselves and let them imitate us as best they can at the time" (p. 224). He concludes that "the study of grammar as such is neither necessary nor sufficient for learning to use a language" (p. 226).

Basing their argument on the way a child learns its native language, Newmark and Reibel (1968) recommend a situational rather than a structural approach to second-language teaching. "The main control the teacher needs to exert over the materials to be studied is that they be graspable as usable items by the learner. The language learning capability of the student will gradually take care of the rest" (p. 161). Another writer, Diller (1975), sees the influence of transformational grammar together with cognitive psychology in a number of new varieties of the direct method.

Others share a totally different view of the implications of transformational theory. Chastain (1971), for example, in a book addressed to prospective and practicing teachers, sees transformational grammar, linked with cognitive theories of learning,⁸ as support for the premise that all learning must be meaningful. The aim of the language teacher is to develop "competence." From this view, a rational, mentalistic approach to second-language learning is derived from transformational grammar, one that still emphasizes the centrality of grammar but attempts to relate it to meaning. Bolinger (1968) arrives at a similar conclusion. He sees the transformational emphasis on competence and on understanding the "how" of language as implying "that the rules of grammar be presented and taught forthright, not left in the care of induction" (p. 39). Whatever the interpretation, transformational grammar has undoubtedly caused language-teaching theorists to rethink their view of language.

⁸ Cognitive theory implies that the L2 teacher should avoid rote learning, use graphic and schematic procedures to clarify relationships, use both spoken and written language to appeal to different senses, and employ the native language, visuals, or demonstrations to build a conceptualization of meaning and form in the L2.

3.3 Reassessment of the Relationship

A number of authors (e.g., Bolinger, 1968, 1972; Rivers, 1968; Roulet, 1972; Pillet, 1974) have endeavored to analyze the benefits and problems in language teaching that have resulted from the impact of linguistics. Some of the failings of audiolingualism can be attributed to failings in structural linguistics. Bolinger's comments on the empirical bias have already been quoted (see the subsection on deep structure and surface structure in Sec. 3.2), and transformational theory has served to some extent to fill the gap with new insights into syntactic structure. Both Roulet and Bolinger praise the results of the audiolingual emphasis on spoken language and authentic current usage: "Both from the standpoint of practical skill and from that of deeper understanding students are better on this score than they were 20 years ago" (Bolinger, 1972, p. 112). The explicit information provided by structural linguistics on the sound system and that provided by phonetics on how to form sounds are also seen as beneficial. Bolinger notes that structuralist theory was weaker on rhythm and intonation and that the results in this area leave more to be desired. Roulet points out that structuralism benefited language teaching by preparing precise and verifiable definitions founded on formal and distributional criteria. The presentation of patterns as unities in drills has served to focus attention on regularities in the language rather than on exceptions, as was customary in language teaching based on traditional grammar. At the same time, Chomsky's critiques, along with those of others, have promoted a much less simplistic view of the nature of language. For example, Lyons (1968, pp. 38-42) points out that there are differences between speech and writing that can be related to their respective media, indicating that "written language cannot be regarded *merely* as the transference of spoken language to another medium" (p. 40).

The problem with pattern-practice drills—the boredom and mindless parroting they have been found to encourage—may be that they, too, are based on a theoretical oversimplification (see, e.g., Bolinger, 1972). In Bolinger's opinion: "Where they were bad was when they fell in the hands of literal-minded teachers and text-book writers who adopted the faith that language was a set of habits *and nothing else*" (p. 113). This view is reaffirmed by others. Allen (1975, p. 71), for example, is one who considers that the present trend away from a view of language acquisition as habit formation towards language as rule-governed behavior does not imply that conscious rule-learning is necessary in all aspects of the L2 syllabus. Both views of language may be relevant in accounting for different aspects of language behavior (see also Carroll, 1971).

Other problems of audiolingualism can be attributed to a more fundamental weakness—the oversimplification and misinterpretation of the relationship between linguistics and language teaching. As Wilkins (1972)

makes clear: "It is a simplistic application of post-Bloomfieldian linguistics that had led in some methods to a too rigid distinction between form and meaning in teaching, the neglect of meaning, the attempt to separate the learning of different linguistic levels and an over-dependence on a behaviourist interpretation of learning." As Wilkins sees it, the "point is not that this is now thought to be 'bad linguistics' (placed in a historical context the principles they followed can be considered well motivated), but that they were applied too unthinkingly to language teaching. The relationship of linguistics to language teaching must be studied afresh" (p. 217). Wilkins sees three ways in which linguistics justifiably influences language teaching: by applications, insights, and implications. The applications of linguistics are limited to the information and analyses provided in language descriptions (see also Spolsky, 1969b). And even these are often not as startlingly new and revealing as teachers have been led to expect.⁹ More important, in Wilkins's view, are the insights provided by linguistics that increase our understanding of the nature of language and language learning. He considers that decisions in language teaching are more soundly made when based on a deeper understanding of what language is. More emphatically, Corder (1973) states that "there can be no systematic improvement in language teaching without reference to the knowledge about language which linguistics gives us" (p. 15). The relationship of "implication," in Wilkins's view, exists most obviously between language teaching and what is known so far about the psychology of language acquisition. It is evident in decisions about methods and techniques: whether, for example, to use translation, explanation, or inductive ways of teaching meaning. Thus we see here, based on experience, a subtle shift from a simple application model to a more sophisticated resource model.

As many commentators have noted (Saporta, 1966; Chastain, 1971; Jarvis, 1972), linguistics has had as much to say about how languages should be taught as about what should be taught. But linguistics is only one of the disciplines that have relevance for language teaching, and it has no direct bearing on methodology. There is no reason to assume that a valid analytical procedure in linguistics should also be valid as a method of language teaching, nor indeed that the linguistic description itself is necessarily "valid" for language-teaching purposes. As Corder (1973) points out:

⁹ As Wilkins puts it (1972): "Linguists themselves must take some of the responsibility for raising the expectations of the language teacher too high. By implying that the preparation of a language teacher was incomplete if he had not studied linguistics, linguists have suggested that the content of language teaching would be changed more radically than could in fact be the case. Many of the facts of language are inescapable whatever one's theoretical viewpoint, and traditional grammarians were as capable of recognizing them as linguists have been. It is hardly surprising, therefore, that there should be very great similarities in the content of language descriptions produced by older and more recent scholars" (p. 223).

Criteria for judging the *validity* of a linguistic description are internal to linguistics. A description is valid if it describes adequately what it purports to describe. . . . Validity is related to goals. It is therefore necessary to make a distinction between validity in respect of theoretical goals and validity for practical or applied purposes. I shall call this latter "validity," *utility*. The test of the utility of a description is, then, whether it promotes more efficient language learning. The proof of the pudding is in the eating. (p. 196)

Pedagogical grammar. When it comes to applying the information and analyses in linguistic descriptions, several scholars have recognized a problem of mediation between linguistic theory and L2 teaching (e.g., Spolsky, 1969b). This problem has led a number of writers to recommend a "pedagogical grammar" as an intermediate step between theoretical analyses and L2 course materials. Allen (1974) defines such a grammar as "a collection of material extracted from one or more scientific grammars and used as a basis for language teaching." The applied linguist must pick and choose among formal statements "in the light of his experience as a teacher, and decide what are pedagogically the most appropriate ways of arranging the information that he derives from scientific grammars" (p. 60). According to Allen: "The contents of a linguistic grammar may suggest, but not dictate, the arrangement of material in a pedagogic grammar, since the essentially pragmatic considerations of frequency, usefulness, relevance to situation, etc., must be built into any grammar which is intended for teaching purposes" (p. 70).

In emphasizing an "eclectic" approach to language descriptions, the advocates of pedagogical grammar, such as Allen (1974), Noblitt (1972), Stern (1974), Wilkins (1972), and Corder (1973), have an interesting example to follow in *A Grammar of Contemporary English* (Quirk, Greenbaum, Leech, & Svartvik, 1970), a comprehensive reference grammar that is based on an extensive survey of English usage conducted by the University of London. The authors of this grammar state that their grammatical framework has drawn on traditional grammar and on insights from several contemporary schools of linguistics. While the grammar is not organized for pedagogic purposes, it represents a remarkable achievement in presenting new information from scientific grammars in a form accessible to the layman.

A number of suggestions have been made for a similarly eclectic approach to the teaching of English grammatical constructions. They represent the kind of analysis that could form part of a pedagogical grammar. An example is Levenston's *Teaching Indirect Object Structures in English—A Case Study in Applied Linguistics* (1973b). He considers that "traditional school grammar, the matrix techniques of tagmemic theory, the rule-ordering of transformational-generative description, the systemic choices of scale-and-category grammar—all these and more can be shown to have their own particular relevance to the language-teaching situation" (p. 2).

Demonstrating the application of these different schools of linguistic theory to teaching indirect object constructions in English, he sums up the utility of a linguistic description under three headings: (a) It supplies *explanations* for grammatical phenomena, (b) it provides guiding principles for *sequencing* grammatical structures in materials, and (c) it supplies guiding principles for *structuring* language teaching materials and tests.

Two recent practical guides to teaching French (Rivers, 1975a) and German (Rivers, Dell'orto, & Dell'orto, 1975) may be regarded as fulfilling many of the functions of a pedagogical grammar. They imply, however, a wider definition which includes insights from psychology and teaching experience.

Just how broadly based the contents of a pedagogical grammar should be is not a question that can be settled easily. Clearly, however, there is much more information that course-writers and teachers need to have about the language they teach than its syntactic and phonological structure: They need to know the full range of meanings that the formal units of grammar can realize, and they need information as to what constitutes appropriate language in context. This question of appropriateness may have to do with the structure of the discourse, but it may also include the social setting. Teachers also need to take "communicative competence" into consideration, a notion wider than linguistic competence that refers to the native speaker's ability to produce utterances appropriate to given social situations. Furthermore, teachers need to have much more detailed information than is presently available on how a second language is acquired and on the nature of language "performance"—including the psychological processes that the speaker and the listener use as they encode and decode chunks of language. Thus it seems that a pedagogical grammar, to be truly useful, should contain insights not only from linguistics (including semantics) but also from sociolinguistics and psycholinguistics. In short, the full range of linguistic research, irrespective of the theoretical affiliation of any particular study, becomes a valuable resource for language teaching. It is up to the applied linguist or language teacher to make discriminating use of this resource.

3.4 New Directions

Structuralism and transformational grammar have already exercised an influence on language teaching. In more recent developments in linguistic theory, that is, in semantics, discourse analysis, and sociolinguistics, we can observe how new linguistic influences begin to take shape. Developments in these fields are no doubt too recent and too tentative for there to have been a significant impact on the L2 classroom, but judging from the books and articles in journals addressed to language teachers (Wilkins, 1972; Corder, 1973; Allen & Corder, 1975), there is a growing awareness of their relevance and a concern to interpret the new insights

for language teachers. Often, theoretical developments reflect concerns that have also arisen in language teaching and thus help to clarify issues by providing a more explicit conceptual framework for discussion. Some indication of the usefulness of these rapidly developing fields may be seen by concentrating on a few examples.

Semantics and the teaching of vocabulary. One of the problem areas in language pedagogy that is traditionally and popularly regarded as a matter of semantics is the teaching of vocabulary. Here, it seems, descriptive and contrastive analyses are sorely needed. In the wake of structural and transformational linguistics, many current texts emphasize grammatical structure and neglect vocabulary. The result is that the advanced L2 learner seems more likely to make mistakes in vocabulary than he is in grammar (e.g., Grauberg, 1971). This problem indicates the need in L2 teaching for a precise analysis of the meaning of words and for effective ways of teaching vocabulary that do not overburden the student with memorizing.

A number of frameworks for analyzing the meaning of words have been proposed by linguists, but descriptive information is still fragmentary, so that it would be premature to talk in terms of "applications" to L2 teaching, though certain "implications" can be discerned.

One such framework is that of *structural semantics* according to which "each language must be thought of as having its own semantic structure, just as it has its own phonological and grammatical structure" (Lyons, 1963, p. 37). It is clear from a study of color terms in different languages, for example, that individual languages divide up the same denotational field in different ways. It must be accepted, therefore, "that the reference of such lexical items as the colour terms is not learnt simply by confrontation of each term in isolation with objects having the 'property' to which the item refers but that they are learnt as a set of items limiting the reference of one another" (p. 54). Semantic structure, then, is defined in terms of certain paradigmatic "sense" relations that hold between items in a particular lexical subsystem. A number of different relations are distinguished, such as incompatibility (*blue, green, yellow*), nongradable antonymy (*male/female, single/married*), gradable antonymy (*good/bad, long/short*), hyponymy or class inclusion (e.g., *scarlet* is a hyponym of *red*), and synonymy. This last relation is seen as context dependent. Thus, *the red one* may be a synonym for *the scarlet one* in a particular context. In Lyons's view, it is unnecessary and undesirable to assume that words have a fully determined meaning apart from the contexts in which they are used, though this assumption is still held by many and has led to great difficulties, for example, in treating synonymy.

Recognition of the importance of structural semantic relationships has a number of implications for teaching an L2 vocabulary. First, it implies that knowing a word, or "lexical competence," involves "knowing" the set

of sense relationships into which it can be entered in a given lexical subsystem as well as knowing what it can refer to in external reality (if it has a referent). Second, it implies that meaning must be taught in context. While an approximate translation of a word may at times be a useful shortcut to understanding a term in a particular context, it cannot by any means be regarded as having taught the student to know that word. Nor can "ostensive" techniques of teaching vocabulary that present a word in association with a referent or visual image be regarded as a sufficient method of teaching meaning, though such techniques may be a useful starting point. Structural semantics would indicate that "for learning to be complete, there must be contrast," since referential boundaries are defined "only when a number of related words have been introduced and associated with relevant objects and images" (Wilkins, 1972, p. 131). From his study of structural semantics, Wilkins concludes that the acquisition of meaning is a gradual process that requires massive exposure to the target language. To the L2 learner this probably means extensive reading in the language or "immersion" in a variety of situations of language use.

For the course writer, the utility of structural semantics lies also in its descriptive potential. A compendium of the various sense relations in a language would form a valuable part of a pedagogical grammar (van Buren, 1975).

The structural analysis of semantics outlined above is not the only type of analysis that has been proposed. Another framework, for example, is that of *componential analysis*. Where a componential approach differs most significantly from structural semantics is in its assumption that underlying the "lexical entries" of a language there are preexisting sets of "atomic concepts" (Katz & Fodor, 1963), which may be considered substantive universals. Working within a transformational generative model of language, Katz and Fodor analyze the different meanings of the word *bachelor* into meaning components such as (human), (male), (animal), and (young).

Despite a number of theoretical drawbacks (Bolinger, 1965; Olson, 1970), one of the problems in second-language teaching that a componential analysis might handle most simply and satisfactorily is the problem of collocational restrictions. A typical language-learning error is to use incompatible combinations of words. The German *begegnen*, "to encounter, meet," for example, may be seen as having a component—"meet by chance"—that the L2 student has not noticed when he uses it wrongly in the sense of "meet by design." A description of both the internal componential structure of the L2 vocabulary and a contrastive analysis between the L1 and the L2 (presuming that the components are in some sense comparable) could therefore be valuable in the preparation of materials or for use in explanation (see, e.g., Barrie, 1972; Connolly, 1973). A danger, however, with componential analysis as a teaching technique in

the classroom seems to be that words may be isolated out of context and may give the impression of their having rigidly constructed, predetermined meanings.

Neither structural semantics nor componential analysis can provide a complete picture of the meaning of words. Social variables are, for example, very relevant in the appropriate use of vocabulary, and some of these will be discussed in a sociolinguistic framework.

Semantics and the teaching of grammar. The common denominator in all current discussion of the teaching of grammar to L2 students is the concern for meaning and meaningful communication. The present efforts in linguistics to integrate a semantic component into a descriptive grammar are thus of a considerable interest to L2 teaching. Since an L2 grammar must somehow be learned, the key question is how to make it more meaningful or "psychologically real" (Brown, 1972) to the student.

From the point of view of L2 pedagogy, however, it seems fair to say that what is psychologically real for the teaching of grammar remains to be determined. It is clear that the communicative competence that L2 teaching is aiming for cannot be equated with a formal description offered by linguistics, since there may be numerous ways of "generating" a grammar for purely descriptive purposes. Nonetheless, the increasingly accepted linguistic view that semantics is the formative or base component of language is more compatible with what psychologists have to say about language acquisition and use (e.g., Macnamara, 1972). Much more experimentation needs to be done with a variety of descriptive linguistic models, however, before any claims to psychological reality can be made for any particular description.

One interesting development that may turn out to be of considerable relevance to teaching grammar is the current concern with *case grammar*. Linguists working in a number of theoretical frameworks have agreed on the need for a set of categories to describe the structure of the "proposition," an underlying semantic unit that is seen, for example, in the grammatical clause in English. Fillmore (1968), Gleason (1968), Halliday (1970), and Chafe (1970) all see the need for a set of "deep structure cases" or "participant roles" such as (in Fillmore's terms) agentive, instrumental, dative, factitive, locative, and objective. According to Halliday (1970), such roles function in the expression of processes (actions, events, states, and relations) to represent "the parts that the various persons, objects or other classes of phenomena may play in the process concerned" (p. 146). In Fillmore's opinion (1968), these roles or case notions comprise a "set of universal, presumably innate, concepts which identify certain types of judgments human beings are capable of making about the events that are going on around them" (p. 24). Leaving aside the question of innateness, it does appear that, as far as is known, a similar set is indeed found in all languages. Case grammar has been seen as having great

descriptive potential for L2 teaching (Brown, 1971; Nilsen, 1971; Aid, 1974). Brown finds it particularly helpful in explaining the use of reflexive verbs in French, while Nilsen and Aid see it as a means whereby materials may be both structurally and situationally oriented.

From philosophy (Austin, 1962; Searle, 1969) comes another new perspective in semantic studies that promises to provide a more functional orientation by classifying sentence types in terms of such notions as promising, requesting, asserting, thanking, advising, warning, greeting, and congratulating. While the traditional L2 student may be familiar with formal sentence types such as imperative, interrogative, negative, etc., such labels have only served to give a rough clue as to their appropriate use. A functional classification could, for example, help to explain that in English the sentence *You may leave now* realizes a command, and *Take care* may be a farewell rather than a warning.

Based on a functional analysis of "speech acts" of this kind, a radical proposal for a totally new approach to the language-teaching syllabus has been made by Wilkins (1973). Criticizing the usual "grammatically based" syllabus in L2 teaching on the grounds that it fails to provide "the necessary conditions for the acquisition of communicative competence" (p. 3), he proposes instead a "notional" syllabus in which the notions that the L2 learner needs or wants to express will determine which forms are most valuable to him. Wilkins's scheme for such a syllabus is first divided into six semantico-grammatical categories, several of which—time, quantity, space—are notions that have traditionally been grouped together in the teaching of grammar. The remainder of the syllabus is concerned with "categories of communicative function. They relate to uses of language where there is at best a very untidy relationship between the function of the utterances and the grammatical categories through which these functions are realized. They are also concerned with expression of the speaker's intentions and attitudes" (pp. 6–7). Included are modality, moral evaluation and discipline, suasion, argument, rational enquiry and exposition, personal emotions, and emotional and interpersonal relations. According to Wilkins, "The value of the notional approach is that it forces one to consider the communicative value of everything that is being taught" (p. 8). The approach is seen as particularly relevant in adult education (Trim, 1971; van Ek, 1975).

A new series of specialized materials for adult students of English, entitled *English in Focus* (Allen & Widdowson, 1974), may be regarded as an application of the notional approach to language teaching. The series provides specific courses for students of English with professional interests in, for example, medicine, mechanical engineering, and physical science. These materials also reflect a concern of language pedagogy (Widdowson, 1973) to utilize recent linguistic and sociolinguistic investigations of "discourse."

Discourse analysis. This rapidly expanding field of study is concerned with "the isolation of linguistic features that differentiate a coherent (connected) sequence of sentences (a text) from an agglomerate of sentences" (Hendricks, 1972, p. 83). While both linguistics and language teachers have tended to concentrate on the sentence as the upper limit of formal grammatical concern, discourse analysis suggests that there are many dependencies between sentences in a text. At the same time, semantic studies have indicated that sentences can rarely be interpreted fully in isolation but most often have to be placed in the wider context of the discourse as a whole.

A number of studies provide evidence of the importance of discourse phenomena to language analysis. Such studies are of great interest to language educators concerned with teaching the appropriate use of sentences. Halliday (1967, 1968, 1970), for example, provides an illuminating analysis of how the "textual function" of language is realized in the structure of the English clause. This function is the realization of language's need to make links with itself and with its surrounding situation. Thus, some aspects of intonation and phrase sequencing in English, which may be regarded as "options" in a sentence grammar, are guided in their textual context by the requirements of "coherent" discourse. One such requirement is the need to distinguish what is "new" information from what is "old" in each "information unit"—usually a clause. Gleason (1968) analyzes some additional aspects of discourse structure; for example, the structural means by which identical participants can be realized in successive clauses and the various ways in which the clauses themselves may be connected in a given language. As Gleason indicates, the possibilities available in these areas may differ greatly from one language to another, as may other features of connected discourse. A contrastive analysis of discourse structure could help to explain the difficulties encountered by L2 students in producing "coherent" discourse. Other enlightening studies of discourse have been undertaken from a sociolinguistic perspective, such as the functional analyses of conversations undertaken by Schegloff and Sacks (1973). These studies suggest some interesting guidelines for the construction of realistic dialogues in L2 teaching materials.

It is clear from these recent developments in linguistic theory that they provide an abundant resource which has the potential for deepening and enriching language teaching. But it is equally clear that this resource is not ready-made for easy application. The same may be said of sociolinguistics, the final development to be considered in this section on the impact of linguistics.

Sociolinguistics. In recent years, a number of authors have underscored the relevance of sociolinguistics, the study of language in its social context, to second-language teaching (Shuy, 1969; Criper & Widdowson,

1975). While transformational generative linguistics has been concerned primarily with the competence of "an ideal speaker-listener, in a completely homogeneous speech community" (Chomsky, 1965, p. 3), sociolinguistics aims at an account of the varied uses of language in its interaction with society.

The relevance of such endeavors for language teaching is plain. L2 teachers have long been aware that mastering the formal system of a language does not automatically lead to the ability to communicate in that language. And yet, the development of "communicative competence," the native speaker's ability to use the language appropriately in context, is increasingly seen as an essential goal of language teaching. Corder (1973) comments: "It is probably unnecessary to point out that when we are teaching a second language we are trying to develop in the learner not just *grammatical competence* in the Chomskyan sense, but *communicative competence*. We are teaching him not only what we call 'the formation rules' of the language, but in addition what Hymes has called 'the speaking rules'" (p. 73). Corder suggests that the inability of students to put to use the language they have learned in class can be explained by referring to the notions of grammatical versus communicative competence. Students may acquire the first but not the second.

Theories of communicative competence, however, are not yet well developed. As in semantic studies, L2 teachers are aware of the potential value of sociolinguistics but lack the descriptive information that would enable them to make applications to L2 teaching. A number of conceptual frameworks have been proposed, however, which provide useful insights into the multifaceted notion of communicative competence.

One such framework is that of the "speech event," which could be a conversation, a sermon, or a recipe for baking a cake. Hymes (1974) discusses a number of dimensions that should be taken into account in what is known as the "ethnography of communication" or the sociolinguistic analysis of the speech event: (a) the "participants" in the event, the senders and receivers of a message, and others involved; (b) the "channels" of communication and how they are used, for example, in speaking, writing, and body movements; (c) the "codes" shared by the participants, such as the language and the geographic or social dialect; (d) the "situation" in which the communication takes place; (e) the "message form" and its "genre"—this term is used in a broader sense than the literary one to include conventional constraints beyond those of the grammar; and (f) the "contents" and attitudes conveyed by the message. Particularly important in a sociolinguistic analysis of the speech event are the immediate social roles and the relative status of the participants involved, where subtle distinctions may be made by means of "keys" (Hymes, 1974) that distinguish, for example, a mock from a serious tone and different levels of formality.

The speech event itself is frequently seen as being composed of a number of "speech acts" defined functionally rather than in terms of linguistic form (see the subsection on semantics and the teaching of grammar in Sec. 3.4). Several functions may be realized simultaneously in a single speech act. Also, broad functional categories are associated with various components of the speech event. Thus, the "aesthetic" function focuses on the form of the message; the "referential" function, on its content; the "phatic" function, on the channel of communication (psychological or physical); and the "metalinguistic" function, on the code. Any utterance may have a number of such focuses.

The systematic analysis of speech events in light of categories and functions such as those described above could provide valuable information for the preparation of materials in L2 teaching. By forcing consideration of all relevant variables, it could lead to more realistic simulations of speech events in the classroom and provide a more adequate basis from which the L2 student can move into real communication with members of the target language and culture. Assuming that the sociolinguist will be able to provide a taxonomy of speech events, a major concern for L2 teaching will be to decide in what kind of events the students will be most likely to participate. This would naturally involve a consideration of the social roles that the student might be called upon to perform and bring about new teaching goals. Are the students studying German for future careers as scientists? Are they learning English for use in business affairs? Do they need general groundwork in the second language and culture from which to move into more specialized fields later on? Sociolinguistics is beginning to provide a better defined basis for selecting materials in language courses for special purposes.

Some writers (Paulston, 1974) have suggested the use of specific exercises based on a sociolinguistic analysis of the speech event for L2 classes. In this way, different speech roles are practiced in events where the participants display different social relationships to one another. In addition, emphasis is placed on teaching the "small talk" of situations "which tend to use phatic language: greetings, introductions, farewells" (p. 355), where nonverbal aspects of communication are important. Other exercises are designed to develop communicative competence in subtleties such as hiding one's feelings in conventional ways in the target language.

The concept of communicative competence is also finding its way into language testing (see Savignon, 1972; Oller, 1975; Spolsky, in press). During the past few years, language testers have been experimenting with different ways of assessing a learner's capacity to communicate and comprehend messages in accordance with the social rules associated with the target language. Thus, instead of testing only the ability to pronounce the language correctly or to use correct grammar, a speaking test may require the student to make a socially appropriate response in a given situation

(Levenston, 1973a; Burstall, Jamieson, Cohen, & Hargreaves, 1974).

While the present discussion has concentrated on the detailed analysis of speech events, another thrust in sociolinguistics has been to analyze the use of language from the broader perspective of the society as a whole. From this "macro-sociolinguistic" perspective (Fishman, 1970, 1972), one starts with a given social unit and analyzes the distribution of languages, dialects (geographical or social), and varieties within it. The notion of "speech community" no longer presupposes homogeneity as in transformational generative theory but is seen rather as a community (large or small) with shared knowledge of rules for communication. Several forms of speech may be in use, and different speech communities may, for example, be described as unilingual, bilingual, or multilingual. There may be different kinds of contacts between speech communities as well as different socially determined attitudes toward the language or variety of language used by certain social groups that lead to language shifts or to language conflicts. The patterns of language use in a society may be relatively stable or unstable at any given point in time.

Planning language policy in a society, including the teaching of second languages, needs to take into account surveys and analyses based on concepts such as those discussed here. In 1970, Fishman commented:

Applied sociolinguistics most often comes into play when decisions concerning language policy have been reached and require implementation. It is to be hoped that as language planning and social planning agencies become more aware of the possible contributions of applied sociolinguistics, they may become more inclined to involve sociolinguists and other language specialists in guiding the decision-making process itself rather than merely in implementing decisions already reached. (p. 108)

Such involvement may already be documented in a few instances. Perhaps the best example in this context is that of the Canadian *Report on the Royal Commission of Bilingualism and Biculturalism* (1967-1970), which included a study of the use of French and English as second languages in Canada. On the basis of a comprehensive analysis of social institutions and existing educational patterns in the various provinces, the report was able to recommend policies for cultivating bilingualism and biculturalism in Canada. Clearly, in any country, administrative decisions as to choice of second language(s) to be taught in an educational setting, and whether it should be obligatory for all students, what goals are relevant, and how the teaching of the second language is to be integrated with other subjects, may be more soundly made when the sociolinguistic implications of one choice as opposed to another are understood. Studies of bilingualism (both as a social and as a psychological phenomenon) undertaken by scholars such as Mackey (1967), Rubin (1968), Lambert (1969), Fishman, and his associates (Fishman, Cooper, & Ma, 1971), and Paulston (1975) are particularly relevant for second-language teaching since all language

teachers may be regarded as "producers of bilinguals" (Fishman, 1966, p. 121).

Sociolinguistics may then, like semantics and discourse analysis, be regarded as having potentially a great deal to contribute to language teaching. While the language education field has demonstrated its interest, in the form of articles and books addressed to language teachers, it is clear that descriptive sociolinguistic studies are as yet generally lacking, and many new insights into language have probably not yet had the chance to be applied in the second-language classroom. The extent to which the promising developments in these fields will benefit language teaching depends partly on how effective the means of mediation are between the language sciences and L2 teaching in the future. This problem will be discussed in Section 6.

4. PSYCHOLOGY AND PSYCHOLINGUISTICS

4.1 *Changing Research Interests*

While the relationship between linguistics and language teaching illustrates a shift from a naive application model to a more sophisticated application model, or to a resource model, the relationship between psychology and language teaching has developed differently. Language teaching has always operated with certain psychological assumptions, and from about the mid-1950s it has used psychology and psycholinguistics as a resource to help find solutions to some of its major problems. In these developments, the relationship has shifted more and more towards a convergent model.

A 1959 conference at UCLA on "Psychological Experiments Related to Second Language Learning" (Pimsleur, 1960) provides a starting point from which to view today's psychological and psycholinguistic research relevant to language teaching. It brought together a team of language teachers, psychologists, and linguists to define their common interests and to discuss needed research in the psychology of second-language learning. The fact that its organizers called it a "first step" in making knowledge considered relevant from psychology and linguistics available to language teachers shows, from a 1970s perspective, how far interdisciplinary cooperation has come in the intervening years. The relevance of psychology, psycholinguistics, and linguistics to language teaching is seldom questioned today.

A comparison of the subjects discussed at the conference with current literature reveals some continuity but also some rather startling changes in emphasis and theoretical stance over the last decade and a half. A number of the 1959 research proposals revealed some underlying behavioristic and neobehavioristic assumptions which could be traced back to the 1954 report on psycholinguistics by Osgood and Sebeok. For example, projects

were outlined to investigate the most efficient types, modes, sequences, and reinforcement patterns of classroom drills without questioning the basic pedagogical utility of drilling. Other projects were designed to investigate such problems as the satiation effect of repetition in different lexical contexts. Since then the validity of the behavioristic model of second-language learning has been seriously questioned, and such studies of specific issues have given way to an emphasis on more fundamental questions such as the role of cognitive processes in language learning.

An assumption apparent throughout the UCLA Conference report was that all students learn languages in essentially the same way, regardless of how well they perform what they learn or their rate of progress. Instead, today, researchers are investigating individual differences in learning styles, strategies, types of associations made, and preferences for particular kinds of language-learning situations. Earlier interest in programmed instruction has waned with the realization that individualization involves varying more than the learning rate, and with the increasing stress on the ability to communicate in real-life situations.

The years following the conference have been productive ones for the relationship between language education and the disciplines of psychology and psycholinguistics. There has been growing interest in second-language learning as an area of psychological and psycholinguistic study. And in spite of the radical changes in learning theory, which have been difficult for educators to accommodate, the language teaching profession has continued to be receptive to psychological interpretations.

4.2 Psychological Variables Related to Second-language Acquisition

One area of psychological investigation into second-language learning that began receiving attention in the mid-1950s, and that is still an important research focus today, involves the definition and measurement of generally applicable "learner factors" related to achievement in language classes. Particular attention was given to applications of the concepts of aptitude, attitude, and motivation. Other variables with apparent psychological implications for second-language learning—such as age, socioeconomic status, sex, and certain personality traits—were also investigated in a variety of learning contexts. Together, these studies led to an appreciation of the complexity of influences on second-language learning. They indicate a growing convergence between theoretical research in psychology and the practical concerns of the language classroom.

The research into learner factors may be illustrated here by a discussion of studies in three areas: language aptitude, attitudes, and the relationship of age to language learning ability.

Language aptitude. Since the 1920s, educational psychology has developed, and refined, the idea of specialized talents, or aptitudes, over and above the general scholastic ability measured by intelligence tests.

Naturally, this concept, along with the psychometric techniques already tested in a number of fields, relates to the age-old experience of language teachers that some students learn a second language more easily than others. In the 1950s, Carroll, as a psychologist interested in language teaching, collaborated with Sapon, a language teacher interested in psychology, in the development of a test to predict the success of adult language learners. Along with their theoretical interest in the definition of language aptitude, the immediate aim was a practical one: to develop a means of identifying those students with a high probability of benefiting from training in an allotted time (Carroll, 1962). This study is an excellent example of convergence between theory and practice.

The Carroll-Sapon Modern Language Aptitude Test (MLAT) was constructed from some fifty experimental tests measuring abilities that had been linked with language learning ability. Using factor analysis, a battery of five subtests was constructed and has been used since 1959. Carroll (1975a) identifies four components of foreign-language aptitude measured by the MLAT: "phonetic coding ability," "grammatical sensitivity," "memory ability," and "inductive learning ability."

An elementary version of the same test (EMLAT), developed subsequently by Carroll and Sapon (1967), measures similar factors in elementary-school-age children.

Pimsleur's (1966) Language Aptitude Battery (PLAB) for high-school students was devised in the early 1960s for Grades 6-12. In addition to measures similar in many ways to those of the MLAT, it includes in its prognostic score a calculation of student grade-point average and a measure of the student's expressed interest in learning a second language. Both measures were found by Pimsleur to add to the predictive value of the test for high-school students in existing secondary-school foreign-language programs.

In these three instruments—the MLAT, EMLAT, and PLAB—the psychological concept of aptitude combines with psychometric techniques and a behavioral analysis of language learning and is put to service in the solution of a practical teaching problem: the prognosis of success in foreign-language learning.

All three aptitude tests are still widely employed today for the selection and placement of students in language courses,¹⁰ and are frequently used as control measures in second-language-acquisition research (e.g., Gardner & Lambert, 1972). They have also given rise to further theoretical discussion about the nature of foreign-language aptitude in the light of recent psycholinguistic theory (Jakobovits, 1970; Neufeld, 1973, 1975; Hancock, 1975).

¹⁰ For example, the language training program of the Canadian government, which is responsible for a vast scheme of second-language teaching to federal civil servants, uses the MLAT in the preliminary selection and placement of students.

Up until now, the pioneer work of Carroll, Sapon, and Pimsleur has not been systematically followed up by investigations seeking to analyze further the abilities measured in the aptitude tests and to consider their relevance to general language-learning processes, individual learning styles and strategies, and given classroom situations. In the meantime, the concept of language aptitude and the practical application of these measures have found widespread recognition in language education and must be taken into account in any future studies on language learning.

Attitudes and motivation. Another current work on characteristics of language learners is in the area of psychosocial research. It was initiated by Lambert and his associates at McGill and is continued today by Gardner and his associates in London, Ontario.¹¹ These studies, carried out in a variety of cultural and school environments, seek to isolate, measure, and describe learner attitudes and motivations that affect the language-learning process. They represent another example of convergence of theory and practice.

While it has also been of theoretical interest in the field of social psychology, this research has a basis in the observations and intuitions of language teachers and others with long and varied experience in language training (Nida, 1957) and can be regarded as a direct response to their questions. It has also been responsive to the need to explain the large proportion of variance in second-language achievement which could not be predicted by such essentially cognitive measures as those of the MLAT (Gardner, 1960; Feenstra, 1967). Pimsleur's inclusion of an "interest" measure in his predictive battery was an early recognition of the importance of affective variables as well as cognitive ones. The study of affect in educational and social psychology, generally, was another initiating factor for this research.

The Gardner-Lambert series of studies (1972) began with the hypothesis, and ultimately led to the description, of two motivational components. The first is defined as an "integrative" orientation in which the language learner is "psychologically prepared to adopt various aspects of behavior which characterize members of another linguistic-cultural group" (p. 3). The second is defined as an "instrumental" orientation in which the main motivation for learning the language is utilitarian, such as career advancement. Early research carried out with Anglo-Canadian and American students of French in Montreal (Gardner & Lambert, 1959, 1972; Gardner, 1960; Lambert, Gardner, Barik, & Tunstall, 1963) and Jewish high-school students learning Hebrew, also in Montreal (Anisfeld

¹¹ Many of the earlier studies are summarized in *Attitudes and Motivation in Second-language Learning* (Gardner & Lambert, 1972). Other researchers who have been active in investigating the role of social attitudes and specific components of motivation in L2 learning include Pimsleur, Stockwell, and Comrey (1962); Pimsleur, Sundland, and McIntyre (1966); Ervin-Tripp (1967); Carroll (1967); Spolsky (1969a); Feenstra (1967); Cowan (1968); and Burstall (1968, 1970, 1975).

& Lambert, 1961), indicated the importance of an integrative motivational component for the successful acquisition of a foreign language in these situations. Such an orientation appeared to be particularly relevant to the achievement of conversational skills, especially in an environment where the L2 community was accessible and where such skills could be improved out of class through interaction with members of the other group. Later studies provided further confirmation of the importance of affective variables in L2 achievement and also demonstrated associations between student attitudes and those of other persons in the social milieu, including teachers, principals, peers, and parents (Feenstra, 1967; Feenstra & Gardner, 1968; Gardner, Taylor, & Feenstra, 1970; Burstall, 1975). Related research in various U.S. communities and in the Philippines led to further understanding of the role of instrumental and integrative orientations in different cultural milieus where the L2 varied in prestige and usefulness. Gardner and Santos (1970) found that in the Philippine context, where the L2—English—is of undisputedly high utilitarian value, "students who are instrumentally oriented and who receive support from their parents for this orientation are more successful in acquiring the second language than students not evidencing this supported type of orientation" (p. 17). They concluded that an instrumental orientation may be quite important and under certain circumstances even more important than an integrative one.

Burstall, in several reports (1968, 1970, Burstall et al., 1974) on findings of the large-scale longitudinal evaluation of the French program in British primary schools—undertaken by the National Foundation for Educational Research—confirmed the close association between certain attitudes of second-language learners and their achievement but questioned the direction of the association. She noted that "early achievement in French affected later attitudes towards learning French and later achievement in French to a significantly greater extent than early attitudes toward learning French affected the subsequent development of either attitudes or achievement" (Burstall et al., 1974, p. 234). She disputed the usefulness of distinguishing between instrumental and integrative orientations, because in the British study the two types of orientation tended to occur together.

Gardner and his collaborators in London, Ontario, recognize today that the motivation of language learners is indeed much more complex than their earlier studies may have suggested (e.g., Gardner, 1973). They have come to the conclusion that while aptitude cannot be much modified by educational measures, attitudes and motivation can be considerably improved by techniques of intervention. As a result of these conclusions an interesting cooperative venture has developed between language teachers and this research group in London, Ontario. Here language teachers have organized "bicultural excursions," "French days," etc., and

researchers, recognizing these as incentive programs, are now measuring the extent to which such programs in fact do improve motivation. At the same time, the cooperation between researchers and educators on this question is seen as an important aspect of such experiments (Gardner, 1973, p. 4). Another experiment along similar lines, aiming to improve the integrative motivation of American high-school French students, was that of Jonas (1971) in which she studied the effects of establishing a long-term contact between her language classes and a group of school children in Upper Volta.

Research evidence, discussed above, on the relevance for student success of the attitudes of important persons in the school and community appears also to have found a response among language educators. While a direct impact of particular studies cannot be claimed, there is increased awareness of, and concern about, affective variables, on the part of teachers and teacher trainers. Proposals for training programs to sensitize teachers to cultural differences and to the importance of interpersonal relations in the language classroom include those of Jakobovits and Gordon (1974), designed to create better understanding by school personnel of student problems in bilingual settings, and those of Hubbard (1975) for sensitivity training in language-teacher education. Curran's counseling approach used in "Community Language Learning" (Curran, 1972; La Forge, 1971; Stevick, 1973) also recognizes positive affect as essential to meaningful communication and thus important in the language learning situation. Interest in affective aspects is further indicated by two trends: (a) the frequent and increasingly analytical references to motivational and attitudinal factors at conferences and in the professional literature (Nelson, 1970; Disick & Barbanel, 1974) and (b) the marked interest of language teachers in "personalization" of language programs in an attempt to make language teaching more responsive to the goals and interests of individual students (Rivers, 1975b).

Optimal age. A third psychological issue in language education that has received considerable attention in postwar years is whether there is an "optimal age" for second-language learning. For educational policy-makers, it raises the related question of whether nonnative language instruction should be introduced at the elementary level. This issue is of particular interest as an illustration of the complex interactions of theory and research with educational policy.

In many parts of the world, educational circumstances dictate that a second language be introduced during the early school years, either because it is needed as a medium of instruction, or because it is an important social skill and the majority of children in the society do not attend school beyond their preteen years. Thus, in these circumstances, early language education is based on practical need rather than on psychological or pedagogical considerations. In many European countries,

for reasons having to do with the organization of the educational system, language study has typically begun around ages 10 to 12 when the children enter secondary school. Circumstances surrounding language education for children in the United States, where foreign languages have traditionally been taught to older adolescents at the high-school level, have been somewhat different. Here the widespread introduction of FLES (Foreign Language in the Elementary School) programs during the 1950s and 1960s was part of a conscious effort to improve the effectiveness of language education. Current theory about language learning was taken into account and radical departures from traditional practice were encouraged in the hope of dramatic improvements in language learning on a massive scale. Programs in Canada in oral French for English-speaking Canadian children at the elementary level and the introduction of primary-school French in Britain were similar responses to public and political pressures in the 1960s.

The belief that young children learn languages more easily and better than adolescents or adults was widespread at the time; it had its basis in commonsense observations of the L2 skills developed, with apparent ease, by immigrant children or children growing up in a multilingual milieu. But what lent it credence for many educators was backing from neuropsychology and psychology for the concept of an "optimal," even a "critical," age for language learning. Penfield's (1953, 1965; Penfield & Roberts, 1959) work with brain-damaged children and adults, from which he drew implications for second-language education, was an important influence. Citing evidence that children who become aphasic through brain damage in one hemisphere can relearn language whereas adults in similar circumstances cannot, he maintained that children in general have a capacity to acquire new speech that adults no longer possess. Thus, during the "optimal age" between 4 and 10 years, a child can learn a number of languages perfectly, but if left until after that age, his "biological clock" determines that he will never totally master a foreign language.

Penfield was not without opponents, who found the evidence insufficient to justify his claims or who cited other reasons against second-language study by young children, such as possible inhibition of mother-tongue development, or identity problems. Others, including child psychologists, gave developmental reasons in favor of an early start to language study (Gesell & Ilg, 1956).¹²

In the face of pressure to "do something" about L2 learning, experimental FLES programs were instituted in many parts of the United States, usually with the feeling of confidence that they were in line with "scientific knowledge" in spite of the scanty evidence available. An attitude typical of proponents of the trend is illustrated by a leader in the FLES movement

¹² The different points of view were discussed in Stern (1967).

who wrote: "Since there is no known limit to the number of languages a small child can learn in a natural cultural setting, one might start with as many languages as there are days in the week. . . . It is my guess, subject to verification in an actual experiment, that the children would understand everything that is said in this limited setting in all five languages by the end of the school year" (Andersson, 1961, pp. 84-85).

In 1962, and again in 1966, recognizing early second-language instruction as an issue of worldwide importance about which all too little was known, the UNESCO Institute for Education in Hamburg held a meeting of experts to review the language programs in various countries for children age 10 and under, to look at the evidence for and against primary-school L2 instructions and to plan a research program on important aspects of the issue (Stern, 1967, 1969). Presentations at these meetings showed that major questions on this issue were still unresolved. The report of the first meeting contained a discussion of research problems in early-childhood language education and suggested questions needing investigation (Carroll, 1967). This was followed by Carroll's preparation of a detailed research guide, which served as the basis of the 1966 meeting (Carroll, 1969a). It discussed research methodology and presented issues relevant to the study of problems in this area all over the world.

As noted in the report of the 1966 meeting, by this time there was no longer any question of the feasibility of teaching second languages to children. But the question of "whether children learn languages better or faster than older persons" (Carroll, 1969b, p. 60) was not yet satisfactorily resolved. While further development of elementary-school language programs—materials, tests, and teacher-training techniques—had taken place, little precise and carefully controlled scientific research into the basic issues had been done since the earlier meeting.

From the vantage point of almost a decade after the second meeting, one is forced to conclude that those making educational policy in this area still have very little to go on from research and must rely, essentially, on their own experience with early language teaching. The carefully drawn guidelines of the 1966 UNESCO meeting were not taken up in any coordinated fashion, while the background discussions and examples presented in the two reports were widely interpreted as endorsements of early-childhood language programs rather than as a call for a concerted international research effort.

The programs instituted during the last two decades show that the introduction of foreign languages in formal settings at an early age is no more a guarantee of successful L2 learning than are traditional high-school or university programs. Studies of particular projects and their effects have in many cases failed to find any overriding psychological or pedagogical reasons for beginning a foreign language in earlier versus later school grades. For example, the recent NFER evaluation of

primary-school French in Britain (Burstall et al., 1974) found that children who started French at age 8 did not, in the long run, reach a significantly higher level of achievement than those who started at age 11. The policy implications of this sobering research report are widely debated in Britain today. Some evidence points to the possibility that young children may have greater facility in acquiring certain language skills than adults—notably a native-like pronunciation (Scovel, 1969; Asher & Garcia, 1969), but adults may also have certain advantages, such as in vocabulary mastery.

One highly successful approach for teaching languages to young children, according to evidence so far (Swain, 1974; Swain & Barik, 1976), is the Canadian French-immersion programs for anglophone primary-school children (see also Sec. 5). But even there the successful language learning cannot be unequivocally attributed to the age of the children because there are other important differences between these programs and the traditional ones given older students. One difference is that anywhere from a half to a full school day is spent in the L2 "environment" with a native speaker as teacher, and, perhaps even more important, the children use the second language functionally rather than studying it formally as a school subject.

Thus the issue of an "optimal" age remains unresolved, providing an example of the often haphazard relationship between research and classroom reality. When education needed quick answers from the language sciences as a basis for new policy, it grasped at what might be regarded as the insufficient and rather speculative evidence available. At this point one may ask whether science has failed education, or if perhaps the hopes of grand solutions were unrealistic in the first place.

4.3 *The Learner and Language-learning Processes*

Theoretical shifts. Although still in its early stages, the most interesting and promising line of psycholinguistic research for second-language teaching consists of a number of recent studies attempting to examine in depth the language-learning process itself. In retrospect, it is astonishing that so little was done earlier in psychology and psycholinguistics to gain a better understanding of the mental processes involved in second-language learning and in the use of an imperfectly known language.

Leading psychologists interested in language learning in the 1950s and 1960s adopted a surprisingly mild attitude vis-à-vis the current psychological interpretation of both first- and second-language learning. They seemed reluctant to criticize the language-teaching profession for working with an oversimplified model, perhaps a reflection of their recognition of the lack of research evidence about anything that might replace it. Finally, it was a leading language-teaching theorist, Rivers, who brought the issues into the open. Her book, *The Psychologist and the Foreign Language Teacher*

(Rivers, 1964), was the first comprehensive, critical analysis of language teaching theory in light of psychological theory and practical teaching experience.¹³ But it was the new view of language acquisition following the introduction in transformational generative grammar that finally led psycholinguists to a thorough reappraisal of the theoretical basis of existing views of language and to a continued search for a more satisfactory interpretation of second-language learning.

From a psycholinguistic perspective, audiolingualism could be challenged on the following points: (a) The audiolingual method assumes that the learner needs a perfect L2 performance model where error making is avoided as much as possible. Appropriate behavior is to be elicited through mimicking the model and then conditioned through the positive reinforcement of all correct responses. Transformational generative rationalists argue that children achieve competence in the target language, from less than perfect models, by abstracting the rule system of the language they are learning. Error making is viewed as a natural part of the learning process, in which the learner makes and tries out hypotheses that he then refines and corrects in response to feedback. (b) Audiolingualism, furthermore, assumes a learning model in which the environment is the main determinant. The developing L2 capability is seen as consisting primarily of learned behavior, explainable in terms of associative learning, transfer, mediation, and conditioning. The learner's L2 behavior is "shaped" over a series of successive approximations so that he moves closer to native-like language behavior. Learning difficulties are viewed as primarily due to inadequacies in the environment, such as poor models or inappropriate reinforcement. The new acquisition theory suggests instead that the learner has an innate capacity, sometimes referred to as the language acquisition device (LAD). The LAD concept attempts to account for the ability of children to learn their native tongue in a natural environment in the absence of perfect models and selective reinforcement based on accuracy. There is some indication that both first- and second-language learners have a "built-in syllabus," or consistent order, in which they acquire grammatical structures (see the following subsection on language-acquisition studies). According to this theory, the learner actively initiates the language-learning process rather than merely responding to stimuli, as suggested by audiolingual theory.

In spite of its being only in the early stages of formation, with many details so far unexplored, the new theory of language acquisition raises serious questions in the minds of language educators. It throws into doubt the assumptions and methods of audiolingualism while offering few guidelines as to what might replace them. Indeed, some question whether

¹³ In the same year, at an international conference in Berlin, Carroll (1966) also took a much more critical view of the psychological theory underlying current views on language learning than he had previously.

language can be taught at all (Dulay & Burt, 1973, p. 257).

The observation that language behavior is "rule governed" led to calls for a greater emphasis on explicit attention to grammar (Diller, 1975). There were proposals to teach language by creating situations where the learner experiences natural language in use, leaving the learners on their own with minimal formal instruction. The L2 program at the University of California in San Diego described by Newmark (1971) is an example of this approach.

The newly perceived value of error making as a clue to the hypotheses of the student, coupled with the observation that L1 learners perfect their language skills without necessarily receiving reinforcement for proper responses or criticism for improper ones is cited as evidence that teachers do not need to correct errors (discussion in Gorbet, 1974). Others, suggesting applications of the research on errors, assume that teachers should indeed correct them and that the research evidence can help the teacher identify and understand the systematic nature of particular learner errors. Dulay and Kiparsky (1972) created *The Gooficon* based on ESL learners' errors to help teachers do this. In the face of confusing interpretations, a common result of the psycholinguistic theoretical controversies is that educators are recommending an eclectic approach to theory (Paulston, 1973) and encouraging teachers to become more actively involved in experimentation (Rivers, 1975b).

Several scholars between 1966 and 1975 sought to reconcile behaviorist and rationalist interpretations of language learning (Anisfeld, 1966; Carroll, 1971, 1974; Rivers, 1972). In a recent article, Carroll tried to pinpoint the merits and shortcomings of both positions: If a cognitive process is to bring about a relatively permanent change in behavior, it must be made automatic, often going out of conscious awareness, through use that is both sufficiently frequent and varied, across enough different conditions. This practice aspect of language acquisition is largely ignored by the rationalist theory. On the other hand, certain cognitive processes are essential to the early stages of learning, in that "the nature of the response to be learned and the conditions under which it is to occur must be perceived by the learner . . . and endowed with some kind of perceptual organization" (Carroll, 1974, p. 127). Learners may act consciously in certain ways to bring about particular consequences. They may also initiate changes in their own behavior to match a model. Learning theorists, such as Skinner, do not take such cognitive processes into account in their theoretical formulations. While it was essentially the radical behaviorism of Skinner that Chomsky attacked, there has been, throughout the history of psychology, a current of theory that assumes the existence of various sorts of cognitive processes and that acknowledges that human beings have "goals, plans, perceptions, expectancies, and ways of organizing experience internally" (Carroll, 1974, p. 129).

On the basis of experimentation and research in learning theory and verbal learning, Carroll (1974) outlines specific guidelines for teachers, learners, and writers of textual materials. For example, in addition to practicing in meaningful contexts, learners should be encouraged to remember meanings conveyed in the foreign language, to remember them while practicing, to try to work out explanations about how the language is put together, and to pay careful attention to explanations given by others. Similar suggestions for those preparing instructional material include the observation that an exact sequencing of content is not essential, but that it is important that new material "fit in with and enhance the learning and memory of old material" (p. 144).

Rivers's efforts over the years to relate psychological theory and research to language teaching practice are illustrated by her suggestions for teaching listening comprehension skills. Based on work in the field of speech perception, in which she readily admits "there are many questions . . . for which at present we can expect no definitive answers" (Rivers, 1971, p. 123), she identifies areas in which there is sufficient research evidence to justify certain teaching activities fairly conclusively. She describes speech perception as a complex, almost simultaneous three-stage process, "which should be kept in mind in the designing of teaching materials" (p. 126). The "sensing" stage of rapid, fleeting impressions, requiring rudimentary segmentation and filtering out of "noise," is followed by a second stage of "identification through segmentation and grouping" (p. 127) and a third involving the rehearsal and recoding of material before its entry into long-term storage. Rivers suggests that listening-comprehension materials be "so constructed that the student has specific practice in the various types of operations he must perform" (p. 129) and describes methods of prolonging the auditory image such as inducing student repetition, rehearsal, and conscious attention to structural segmentation and teaching students how to recode. Parallel suggestions follow for teaching reading.

In the face of theoretical confusion, it is valuable to have guidance in specific applications from scholars such as Carroll and Rivers, who from different vantage points have undertaken the task of mediating between psychology, psycholinguistics, and language teaching. They have shown that the application model, judiciously employed, can be fruitful; but in the long run, it is not enough. This realization has led to new directions in research which we will now consider.

Language-acquisition studies. An important result of the current theoretical controversies in psycholinguistics is a series of studies investigating language learning directly and empirically, through the observation of learner behavior and the analysis of his linguistic output. Psycholinguists involved in this research see their first task as the development of a detailed model of language acquisition based on the

observation of children learning their native language. What those working on this task ultimately seek is an understanding of language competence and its development derived from an analysis of performance data. By studying the acquisition of a small number of morphemes and syntactic structures, they are trying to establish whether there are regularities according to which children adopt particular elements in a given mother tongue (Brown, 1973; de Villiers & de Villiers, 1973).

A second major issue is to define the relationship between first- and second-language acquisition. Are the processes and sequences essentially the same? What differences are there, if any? How important is the makeup of the linguistic "input" or, as Corder would add, "intake" by the learner in determining the acquisition sequence? This area of concern has great potential relevance to the language teacher because, if a consistent "built-in syllabus" is determined, "we may be able to allow the learner's innate strategies to dictate our practice and determine our syllabus; we may learn to adapt ourselves to *his* needs rather than impose upon him *our* preconception of *how* he ought to learn, *what* he ought to learn and *when* he ought to learn it" (Corder, 1967/1974, p. 27). On this question, researchers have again concentrated on plotting sequences in the acquisition of small sets of morphemes or particular syntactic structures to determine (a) whether there is a "natural" sequence for learners of a particular L2 that is independent of the mother tongue they speak and, if so, (b) whether such a sequence matches what has so far been sketched out for L1 learners.

The studies carried out so far provide conflicting evidence on this question. Some have uncovered similarities in the acquisition sequences of L1 and L2 learners in a given language (Bailey, Madden, & Krashen, 1974; Ervin-Tripp, 1974; Ravem, 1974) or for learners with different mother tongues learning the same L2 (Dulay & Burt, 1974b; Khampang, 1974). Others question the parallelism between acquisition sequences for L1 and L2 learners (Dulay & Burt, 1974c; Hakuta, 1974; Hansen-Bede, 1975).

Another unresolved question is the extent of mother-tongue interference for L2 learners or of cross-interference of the two languages of child bilinguals (Dulay & Burt, 1974a; Mougeon & Hébrard, 1975).

At the moment, it is not clear that these studies will yield practical guidelines for pedagogy, but a small number of investigators are now attempting to look beyond the details of these investigations toward a more comprehensive view of second-language learning (Hatch, 1974; Schumann, 1975).

Error analysis and the study of interlanguage. In addition to second-language acquisition studies patterned after L1 research and mainly concerned with the order in which certain new linguistic elements are acquired, a related research approach also involves the analysis of the

learner's linguistic product. The focus here is on his "errors," that is, his systematic digressions from a native language model or his total linguistic output at a given stage, which have been called "interlanguage" (Selinker, 1972; Corder, 1975).

While the analysis of student errors by language teachers and writers of textbooks has a long tradition in language pedagogy, error analysis in its present form is partly an offshoot of contrastive analysis. When it became apparent that learner errors predicted through contrastive analysis of the languages often did not occur and that many learner errors could not be accounted for on the basis of differences between the L1 and L2, attention turned to the systematic analysis over time of the errors themselves (Richards, 1974). This, in turn, led to the positing of developmental stages in learner language and to the idea of studying learner "transitional competence" (Corder, 1967/1974), "interlanguage" (Selinker, 1972), or "approximative systems" (Nemser, 1971), at various points in the learning process.

Error and interlanguage analysis, while they may lead to information on acquisition sequences, are primarily concerned with the learner's overall cognitive approaches to the material being learned. They analyze as a system the whole of the learner's errors or his total linguistic output. Such research has led to the postulation of cognitive acquisition strategies and uses of the L2 such as simplification, transfer from the L1, and transfer from information received in formal L2 training (Selinker, 1972).¹⁴

As for the relevance of these investigations for language pedagogy, it should be noted that, so far, most of the empirical investigations have studied unsophisticated learners of a second language in a natural environment, that is, immigrant children or adults who make no special effort to study the language. However, in a literate society, the notion of a natural or free language learner may be fictitious because, in effect, the learner lives in an educative society and may be more or less his own teacher even if he has no formal instruction. There is, therefore, an equally interesting task and, from a practical point of view, an even more important one: to study second-language output in a more or less formal teaching environment.

Individual differences in language-learning processes. Another research approach focuses on the highly successful language learner and

¹⁴ This research is leading to an increasingly complex model of language development and use. Aspects that have been suggested as important for a thorough exploration of learner language-acquisition and -performance include: (a) the appropriateness of the learner's utterances vis-à-vis the sociolinguistic context; (b) how close the utterance approximates what the learner wanted to say, influenced by the learner's evaluation of his competence in the L2 (Varadi, 1973); (c) the need to elicit representative structures, phonetic and lexical content—based to some extent on contrastive analysis—to complement the analysis of spontaneous texts (Kellerman, 1974; Schachter, 1974); and (d) the utilization of metalinguistic data to benefit from learner insights into the learning process (Corder, 1975).

attempts to identify common and individual characteristics in the language-learning process. Both cognitive and noncognitive aspects have been studied, and attempts have been made to find out whether good learners have specific learning styles or strategies. Also of interest is the question of the extent to which different learners respond to variations in the language-learning situation.

These studies gather data via behavioral observation procedures, testing, and in-depth interviews with both language learners and teachers. A number of attitudinal components, strategies, and techniques of successful classroom learners are described: (a) a dynamic approach by the learner in seeking out opportunities to use the L2; (b) self-monitoring of output and both the solicitation of and sensitivity to feedback; (c) role playing when using the language in the classroom; (d) the voluntary use of monologues in the L2; (e) the seeking of L2 input in different modalities—spoken and written; and (f) conscious association-making of different kinds (Naiman, Frohlich, & Stern, 1975; Naiman, Frohlich, & Todesco, 1975; Rubin, 1975; Stern, 1975; Wesche, 1975). These studies suggest the possibility of identifying cognitive characteristics, strategies, and techniques that are more effective in some learning situations than others. Follow-up research should bring more precise knowledge about how to "match" certain kinds of learners with different learning contexts for optimum results. Furthermore, it opens the possibility of experimenting with teaching specific learning techniques to improve learning efficiency.

4.4 *Conclusions*

While language educators continue to look to psychology for insights and implications for language-teaching theory—albeit more warily than in the early days of audiolingualism—psychologists and psycholinguists are beginning to develop a psychology of language learning based on detailed observation and experimentation in this area of human behavior rather than on extrapolations from work done elsewhere. Hopefully, it will form a basis for an adequate theory of language learning that will eventually encompass first- and second-language acquisition, language pathology, and the conditions and consequences of bilingualism in all its variety (Stern, 1970, pp. 64–65).

Overall, we are coming to see second-language learning and the failure to learn more and more as the result of an interplay of environmental factors, such as the teaching situation, with learner factors—such as maturity level, aptitude, personality characteristics, motivation, previous language learning experience, and learning strategies. A few years ago, McNeill (1966) described first-language learning as "creation of language," in order to point out that the L1 learner is very active in the language acquisition process. In the same way, L2 learning is no longer

seen as a passive, purely receptive, or reactive process, but rather as an active and dynamic process of language creation. The trouble with second-language learning is of course that in so many cases this creative process becomes arrested at a low level of competence. What researchers are increasingly, and rightly, trying to do today is to understand better the dynamics of the learning process and to discover and examine the causative factors that prevent the interlanguage of the learner from being constantly revised, as is the interlanguage for the first language. Once these factors are better understood, it should be possible to devise means for more effective intervention by educators for the benefit of the learner than can be provided at present. This approach to a differentiated understanding of the process of second-language learning and of the failure to learn appears to many of us today as the most important task that psycholinguistic research can perform over the next decade in the interests of language teaching.

Psychology and psycholinguistics can thus be seen as another resource for language teaching to draw on. They contribute concepts, analogies, and knowledge that help in interpreting second-language learning in psychological terms. But, in addition, a reverse process has occurred: Second-language learning is being studied directly and is contributing to a psychology of language learning.

5. EDUCATIONAL RESEARCH

While this chapter is primarily concerned with the impact of the language sciences on language teaching, at least some mention should be made of a series of investigations undertaken within the framework of educational institutions. Such studies have attempted to solve pedagogical issues in language teaching using the experimental methods of educational research. Most of the studies, however, have either explicitly or implicitly also been concerned with psychological or psycholinguistic issues. One example is the famous Scherer-Wertheimer experiment (1964), in which two approaches to language courses in a university were studied: those with an audiolingual introduction and those without. From another point of view, it was a study of the psychological characteristics of language learning under two different sets of conditions.

This group of studies is significant as evidence of the increasing tendency since the early 1960s for language teachers and administrators to adopt a research approach to issues in second-language education. But the studies have also brought with them the realization that research provides only part of the answer.

The earliest studies go back to the beginnings of educational research in the interwar years when Thorndike (1928) made a case for adult learning on the basis of an experimental study that claimed to show that adults

were more effective language learners than children. Thorndike (1931) also investigated experimentally another language-teaching question—whether the teaching of grammar rules is a help or a hindrance. Apart from a few further experimental studies on language learning between the 1930s and the 1950s (Simmins, 1930; Agard & Dunkel, 1948), the most determined efforts to resolve language issues by experimental methods have been made between 1960 and the present. Most of these studies compare two different “methods” of teaching or two different approaches to language problems. In almost all of them, the investigators have jumped straight into the hottest areas of controversy and have inevitably found themselves embroiled in the arguments.

Four well-known groups of studies have attempted to resolve the question of whether the audiolingual (or functional skills) approach to language teaching is more effective than the traditional grammar-translation (or cognitive code) approach. These studies are the Scherer-Wertheimer investigation (1964) at the University of Colorado, the GUME Project in Sweden (Levin, 1972), the Pennsylvania Project (Smith, 1970), and the inquiries undertaken by Chastain (Chastain & Woerdehoff, 1968). The Pennsylvania Project, together with the earlier Keating Report (1963), also tried to gauge the effect of the language laboratory on language learning. One of the most sustained research efforts in language education was recently completed in Britain by the National Foundation for Educational Research (Burstall et al., 1974). It consisted of a 10-year study to find out whether earlier language instruction beginning at Grade 3 is more effective if it is begun a few years later. A radical approach to experimentation is provided by a number of current Canadian studies in “bilingual education” or “immersion,” mainly carried out in Montreal, Ottawa, and Toronto. These studies compare the effect of “immersion” with other forms of second-language learning (Swain, 1974; Swain & Barik, 1976; Stern, Swain, McLean, Friedman, Harley, & Lapkin, 1976). They were pioneered by research initiated by Lambert in the mid-1960s (Lambert & Tucker, 1972). The initiative for all these Canadian experiments came from parents and educators—researchers were brought in to advise and evaluate. Lastly, two international studies initiated by an international research team and undertaken in the context of a vast program of educational comparison, in over 20 countries, have investigated the teaching of French and English as second languages in 8 and 10 countries, respectively (Lewis & Massad, 1975; Carroll, 1975b). In these studies the achievement of students in schools was correlated with a number of current issues in second-language teaching—starting age, length of time devoted to language study, methods of instruction, technological aids available, and various environmental factors.

In addition to these “global” studies in educational settings, several small-scale experimental investigations have been undertaken in which

specific aspects of teaching and learning phonology, vocabulary, or grammar have been studied (McKinnon, 1965; Crothers & Suppes, 1967; Carton, 1971). Most of them have been concerned with the distinction expressed by the Swedish GUME Project between "explicit" and "implicit" language teaching, that is, teaching with, or without, explanation of language rules.

Most of the global studies, except for the immersion projects, have by and large been unable to provide any clear evidence that the innovations—the audiolingual approach, the use of the language laboratory, or the early start—are more effective than traditional methods. This of course, has upset some of the innovators who have attacked and tried to discredit the researchers or the method of research. The research on immersion programs has had a different fate: It reports that immersion is largely more successful than the conventional approach. This finding, pleasing to the innovator, is being attacked by those who have misgivings about immersion as a valid educational alternative.

In retrospect, we can see that the experimental study of language teaching has run into several difficulties. If an experiment is undertaken in a natural school setting, as was the Pennsylvania Project, it is difficult to control or isolate the distinguishing features of the comparison groups. On the other hand, if research is undertaken under laboratory conditions, as were the Crothers and Suppes experiments, it can be criticized for being too remote from the reality of language teaching in classrooms.

A second problem has been the lack of coherence and continuity among the various research efforts. Investigators have studied isolated problems of language teaching and learning without having a theoretical framework in which all the different factors that play a role in the total enterprise are given due consideration. Issues, such as the relative merits of different methods, may be taken up by a project and dropped once the project is finished. To our knowledge, few of the major research projects fit into a coordinated program of studies based on a theoretical framework of pedagogical inquiry.

Third, because of this lack of theoretical framework, the problems under investigation in many of the major or minor experimental studies have often been insufficiently conceptualized, and a number of studies have operated with fairly provisional categories of comparison. Neither from a theoretical nor from a practical point of view have such pairs of concepts as immersion versus nonimmersion, audiolingual versus grammar-translation, laboratory versus no laboratory, been clearly enough distinguished.

If experimental research is to contribute to the improvement of language teaching, it is today more and more clearly recognized that certain conditions have to be fulfilled:

1. Experiments should fit into a program of investigations, whether psychological, linguistic, educational, or multidisciplinary in orientation.
2. The conceptual and theoretical basis of the problem or argument must be clarified.¹⁵ While one might *begin* an enquiry by setting out from common distinguishing labels such as "traditional method" and "direct method," for the purposes of the research, these distinctions must first be analyzed theoretically and then investigated empirically: Do the labels mean anything in practical pedagogical terms? In other words, the theoretical distinction must be defined, and refined, and backed by classroom observation and other techniques for documenting the instructional process.¹⁶
3. Studies of teaching should not only be related to the outcome, the proficiency shown by the learner, but to the learning processes that are to be stimulated by the particular form of teaching.

6. MECHANISMS OF INTERACTION

The preceding review of the impact of research on language teaching provides evidence of the vitality and vigor of the relationship and, at the same time, of the problematic nature of their points of contact. In this section, some of the mechanisms that have evolved in developing and maintaining these contacts will be briefly analyzed.

Various kinds of mechanisms can be observed: (a) mechanisms for initiating and carrying out research and development; (b) those for mediation and information exchange; and (c) those for teacher training.

6.1 *Mechanisms for Initiating and Carrying out Research and Development*

In the past (and to a certain extent even today), language pedagogy tried to resolve its problems by ad hoc methods that ultimately led to skepticism about the possibility of any systematic improvement. An eclectic language-teaching theory evolved which allowed languages to hold their own among curriculum activities in schools and universities. Yet, language teaching was constantly being criticized by the public, the students, and to some extent the teachers themselves. Language teachers

¹⁵ The need for a theoretical basis in studies on bilingual education, for example, is stressed by Paulston (1975).

¹⁶ A good example of what is needed is a study by Cooke (1974), carried out between 1968 and 1974. Cooke focused on one aspect of the method argument, that is, the question of the role of linguistic explanation in L2 teaching. He critically reviewed the research literature, developed a model of explanation, and considered how it might fit into the learning process. He then studied the role of different kinds of explanation in a concrete teaching and learning situation.

defended themselves against criticism for the low level of proficiency attained by their students by advocating a modest realism as the only policy possible for the language classroom. However, the language-learning needs as viewed by governments, educators, and other social or economic agencies were often so pressing that this modest eclectic realism, to which teachers had accustomed themselves, was unacceptable. It is because of these sociopolitical pressures that a systematic research approach has spread since the 1940s.

Military and political needs, backed by public funding, initiated the wartime efforts in language research and training in the early 1940s. A similar initiative lay behind the National Defense Education Act (NDEA).

NDEA, 1958. The National Defense Education Act was passed in the aftermath of the launching of Sputnik in 1957 and was part of a vigorous reaction by the United States government and educational systems to the supposed competitive threat of the Soviet Union in this area. A population competent in foreign languages and knowledgeable about foreign cultures was seen as essential to the national interest in politics, business, and defense (Parker, 1961). The U.S. Office of Education was prepared to make a substantial financial commitment toward this goal, and, in fact, some \$135 million was spent under the Act during the first five years.

Under Title VI of the Act, the Language Development Program provided funds for the development of foreign-language teaching materials, training of foreign-language teachers, and fundamental research in theory and methods of foreign-language teaching. A number of surveys were sponsored to assess such areas as foreign-language enrollments, numbers and qualifications of teachers available at different levels and for different languages, teacher training requirements, available instructional materials, and current instructional practices.

A major area of emphasis was the preparation of curriculum materials and qualified teachers in the "neglected" or "uncommonly taught" languages. Over a hundred contracts were let for the preparation of such materials—including textbooks, tests, dictionaries, and grammars. University "language and area centers" funded by the NDEA, together with a fellowship program for potential university language teachers, and summer institutes for secondary- and elementary-school language teachers all contributed to this aim, as well as to the development of programs and expertise in the commonly taught languages.

Teams were set up to develop audiolingual teaching materials in a number of languages. The resulting A-LM multilevel curriculum sequences in French, German, Italian, Spanish, and Russian set a pattern for the development of commercial packages and were also noteworthy in that they brought together linguists, psychologists, and language teachers in a joint effort. They provided tape recordings and visual aids that were designed for use in the language laboratory and the classroom. Audiolingualism spread not only through these materials but through hundreds of

summer institutes that sought to upgrade both the language skills and methodological skills of the thousands of language teachers who participated in them. Linguistic theory and cultural courses were regularly taught at these institutes.

Funds from the NDEA also supported research, materials and program development, and teacher training for foreign-language programs in the elementary school (FLES). In addition, a number of "pure" psycholinguistic studies related to language teaching and learning were funded, including some of those previously referred to as proposals of the 1959 UCLA Conference—itsself funded by the NDEA.

Although NDEA funds began to dry up at the end of the 1960s, and few of the programs continue in the same form today, the effect of the NDEA on language research and pedagogy has been enormous not only in the United States but throughout the world.

Other initiatives. Other funding bodies and governments in a number of countries initiated language research and training programs in the postwar decades. For example, the Ford Foundation has consistently promoted language research and development since the early 1950s (Fox & Skolnick, 1975). In France, it was the Ministry of National Education that established the Commission to develop *Le Français Fondamental* in 1951. And in Britain, it was the Ministry of Education, together with the privately endowed Nuffield Foundation, that sponsored long-term research on the teaching of French in the primary school (Stern, 1969).

Research initiatives, of course, do not come exclusively from social and political pressures. As this chapter has made clear, much relevant research has emanated from theoretical studies in the language sciences undertaken in the furtherance of knowledge rather than with the needs of language teaching in view. Such initiative has arisen primarily in the new departments of linguistics in universities and from disciplinary studies that were undertaken in the postwar decades. Curiously, the traditional language departments, with few exceptions, have taken no part in this research in theoretical linguistics and contemporary languages. The research in the universities has, of course, not all been in the purely theoretical sphere. Some university departments, such as the psychology department at McGill University, have undertaken research with a strongly applied orientation (Lambert & Tucker, 1972).

Few research projects have been initiated directly in the classroom. The study by Scherer and Wertheimer (1964) and the Pennsylvania Project (Smith, 1970), which were prompted by the polemics on language teaching methods, are notable exceptions.

6.2 *Mechanisms of Mediation and Information Exchange*

A major characteristic of the developing relationship between the language sciences and language teaching over the past thirty years has been the growth of a number of mechanisms that act as intermediaries. Their

importance has steadily increased.

In the 1940s and 1950s, when the links between the language sciences and language teaching were first forged, the task of applying linguistics was envisaged as a more straightforward and linear undertaking than subsequently proved the case. With growing awareness of the complexities of the relationship, much attention has recently been paid to the mechanisms that mediate between language research and language teaching. As an intermediary discipline, applied linguistics, or educational linguistics, has developed. As institutions, various language centers or institutes fulfilled this function. The need for synthesis and interpretation of what theoretical and descriptive linguistics has offered is also illustrated by the budding concept of a pedagogical grammar. In addition, a whole host of organizations—associations, conferences, publications, and language information centers—have taken up the task of mediation and information exchange.

Because of the lack of a specific locus for initiating research on practical language problems, a number of language centers and similar institutions were set up by governments and funding bodies to promote needed language research. Instigating research has, for example, been one of the prime functions of the Center for Applied Linguistics (CAL), established in 1959 in Washington, D.C., by the Modern Language Association with support from the Ford Foundation. Other centers with a strong research orientation can be found in various parts of the world.

Training programs in linguistics for foreign students in American universities and the establishment of language centers in many parts of the Third World were related to the postwar international commitments of the United States and of its own heightened emphasis on ESL and foreign-language education. Organizations specializing in international development, most notably the Ford Foundation, undertook a number of language development projects in Asia, Africa, and Latin America starting in the late 1950s (Fox, 1975). The early emphasis was on establishing centers and providing resources for teaching English, particularly in countries where English was a major language of instruction. American university professors and often their home institutions became involved in this ESL movement abroad, largely through the Fulbright exchange program. It was natural that the prevailing methodology and curriculum materials in North America be introduced abroad through such contacts. Programs of the Center for Applied Linguistics led to increased cooperation among the various governmental, private, and academic institutions involved in ESL abroad and to a greatly enhanced resource base for their work in terms of trained specialists, curriculum materials, and basic research (Fox, 1975). Although emphasis in a number of the projects has since shifted from ESL to bilingual education and native-language instruction (e.g., the Language Study Center of the Philippine Normal College),

and from specialized, often elite, adult groups to work within school systems serving the public at large, the methodological and linguistic influences of the earlier projects continue to play a role.

Applied linguistics. Applied linguistics first came into its own in the late 1950s as a subfield of linguistics, its purpose being to apply linguistic theory to such areas as first- and second-language teaching, language pathologies, communications technology, and machine translation. It has come to play an increasingly important role in second-language education, basing its underlying theory on psycholinguistics, sociolinguistics, and education as much as on the parent field of linguistics. Examples of recent major publications in the field of applied linguistics are Corder's (1973) *Introducing Applied Linguistics* and *The Edinburgh Course in Applied Linguistics* (Allen & Corder, 1973, 1974, 1975).

Another index to the growth and development of applied linguistics has been the establishment of active international and national professional organizations in this field. The *Association Internationale de Linguistique Appliquée* (AILA) was founded in 1964 by an international group of scholars. One of its major activities is the sponsorship of an international, multilingual congress every three years. The fourth congress was held in Stuttgart in August 1975. National associations affiliated with AILA were founded in Britain in 1967, Germany in 1968, and Canada in 1969. In the United States several different organizations carry out similar functions. For many years the Linguistic Society of America was an active intermediary between linguistic theory and language education, followed by the Modern Language Association. At present the American Council on the Teaching of Foreign Languages and the TESOL Association are very active in this area.

One of the major journals dedicated to applied linguistics is *IRAL*, the *International Review of Applied Linguistics in Language Teaching*, a quarterly which publishes both theoretical and practical articles and reviews in English, French, and German. A number of journals for language teachers also publish articles and reviews in applied linguistics.

Language centers. Since the early 1960s there has been a remarkable proliferation of institutions, known collectively as language centers, charged with a variety of activities related to language teaching and learning. Often connected with universities, they are in some cases independent or sponsored by educational systems. Among those that became prominent in the 1960s, the Center for Applied Linguistics is outstanding. Others are the Center for Research on Language and Language Behavior at the University of Michigan, the International Center for Research on Bilingualism at Laval University in Quebec, the Modern Language Centre at the Ontario Institute for Studies in Education in Toronto, and the Centre for Information on Language Teaching (CILT) in London. Other centers that have promoted research and teaching of particular languages

include the English Language Institute at the University of Michigan, the Centre de Recherche et d'Etude pour la Diffusion du Français (CREDIF) in France, the Regional English Language Centre at Singapore (RELC), and the English Teaching Information Centre (ETIC) of the British Council in London.

Language centers are task oriented by nature and are often the vehicles through which applied linguists carry out their activities. Such centers generally have an interdisciplinary orientation and may combine several functions besides the research function already mentioned. These include second-language training, language teacher education, curriculum development and evaluation, and documentation and dissemination. In all these ways, these centers fulfill their role as mediators.

Professional associations, conferences, and journals. A number of professional associations of foreign-language teachers promote active interchange between their memberships and specialists in the language sciences through conferences, newsletters, and journals. The emphasis on foreign languages in the United States during the 1960s brought increased participation by teachers in such organizations, and the quality of related professional journals steadily improved. *TESOL Quarterly* is an example of the interdisciplinary mediating role of such journals in its specific call for, among other things, articles dealing with "implications and applications of research from related fields, such as anthropology, communications, education, linguistics, psychology, sociology" (*TESOL Quarterly*, March 1975).

The American Council on the Teaching of Foreign Languages (ACTFL), founded in 1967 by the Modern Language Association in collaboration with the National Federation of Modern Language Teachers Associations, has become a centralizing body for the field of language pedagogy. Its main activities include an annual meeting, a newsletter, and a quarterly journal, *Foreign Language Annals*, that publishes articles, research reports, an international bibliography, and the annual *ACTFL Review of Foreign Language Education*.

The National Federation of Modern Language Teachers Associations also publishes a journal of wide interest to teachers of different languages, the *Modern Language Journal*, which frequently contains articles on topics in applied linguistics and psycholinguistics.

In addition to professional conferences sponsored by national and local organizations for language teachers, several regular regional conferences and their published reports have become important forums in the United States for discussing theoretical and practical issues in language teaching. Some of the most important include the annual Northeast Conference, the Kentucky Foreign Language Conference, the Southern Conference, and the Central States Conference. The 1968 Northeast Conference, for example, was devoted to "Foreign Language Learning: Research and Development."

Documentary sources. An astonishing number of documentary sources have been established that strengthen the process of mediation. Prominent among these is the ERIC Clearinghouse on Languages and Linguistics, operated by the Center for Applied Linguistics, which collects, abstracts, indexes, and disseminates information on a wide range of topics through the sale of microfiches. Similar functions for Britain and Europe are fulfilled by the Centre for Information on Language Teaching, which is notable for its *Language Teaching Bibliography* and its quarterly *Language Teaching and Linguistics: Abstracts*. A number of excellent research reviews ranging from Buchanan and MacPhee (1928) to Birkmaier (1973) also provide documentation on language-teaching research.

A recent guide to the kinds of mediation resources referred to above is provided by Yaldan (1976).

6.3 Teacher Training

As has already been pointed out, university departments of modern languages, in which future second-language teachers receive their basic education, have so far provided little contact with the language sciences. Consequently, as a profession, language teachers have not been prepared for, or responsive to, disciplined inquiries of a linguistic, psycholinguistic, or sociolinguistic nature. Most of them have not come into contact with the theories and research of the kind discussed in this case study until *after* having left the university. Some receive a first hint of them during professional preservice training. Others first become acquainted with them in graduate programs, in-service training, and through the professional literature and conferences. Therefore, the role of handbooks for foreign-language teachers has become exceptionally important in interpreting the language sciences for practitioners. Since the early 1960s, a long line of such handbooks has appeared, among them some outstandingly good ones; for example, Brooks's (1960) *Language and Language Learning* was one of the first to attempt to relate linguistic and psychological theory to classroom practice. Among other influential handbooks fulfilling a similar function are a number of books by Rivers (1964, 1968, 1972). These books have been much appreciated by teachers because of the author's capacity to temper theoretical insights with wisdom gained from practical experience.¹⁷

7. REVIEW AND ASSESSMENT

The impact of scientific research on an educational activity can partly be attributed to the influence of great men of science. Such men leave

¹⁷ Among numerous other such handbooks, mention should be made of those by Halliday, McIntosh, and Strevens (1964), Lado (1964), Mackey (1965), Valdman (1966), Gritner (1969), Politzer (1970), and Chastain (1971).

their mark on a discipline—for example, psychology—and, as a result of the importance of that discipline for education, may transform educational theory or practice in a decisive way. After they have exercised their influence, it is hardly possible any longer to think about that aspect of education in quite the way one had before. Men like Dewey, Thorndike, Freud, and Piaget have had this type of deep-going influence on educational thought in the 20th century.

Sometimes it is not so much a person as a dominant scientific concept, technique, or theory that has made an impact on education. Such is the case with the concept of intelligence, the notion of the Intelligence Quotient, and, more generally, the mental testing movement. These have influenced educational assessment, the definition of educational objectives, the development of an instructional technology, and the organization of instructional groups.

Then again, there is a third way in which scientific developments can influence educational practice, and it is this kind of influence that is illustrated by the present case study of the impact of research on language teaching. There is no one scientist nor one scientific concept that can be singled out as having exercised a *decisive* influence on language teaching. Rather, language pedagogy has experienced the impact of research in a much more general way. Over the past hundred years and particularly in recent decades, the developing language sciences have made an immense impact on all phases of language teaching. It is no longer possible to think or talk about language pedagogy or to make decisions in this field without taking note of their pervasive influence. This is not to say that there are no great figures whose contribution can be identified. Among earlier linguists, Sweet, Palmer, and Jespersen stand out as pioneers; in more recent decades, Bloomfield, Fries, and Chomsky among linguists, and Carroll among psychologists, can be singled out in this way. Nor can it be denied that specific language-related theories, techniques, and concepts have made their way into language teaching: the phonetic alphabet of the 1880s, the concept of linguistic structure in the 1940s, the notions of "habit" and "skill" in the 1950s, "competence," "performance," and "creativity" in the 1960s, and more recently "communication." Throughout the hundred-year period reviewed in this case study, however, it is much more the influence of the language sciences as a whole than the impact of a single figure or concept that has been felt in language teaching.

7.1 *The Impact of Linguistic Theory*

In our opinion, theoretical linguistics has benefited language teaching by creating among language teachers an awareness of language beyond the particular language taught. There is today a much clearer understanding than there used to be of what it is we are doing when we teach or test a language. As we observed in the brief historical review, there were

of course exceptions in the earlier stages of language teaching theory: Gouin and Palmer, for example, had highly developed theories of language well before linguistics had evolved as a modern science. But their theories were of an ad hoc nature. In the modern era, language teachers have a body of theory to refer to. This is a great asset. On the whole, language pedagogy has attempted to absorb the developing linguistic theories into its own language-teaching theories. Consequently, in more recent writings on language teaching and in course materials, we have observed the shift from traditional grammar to structural linguistics, then more recently to transformational grammar, and, in present-day thought, to the rising interest in semantics and sociolinguistics.

The constant changes in theoretical interests, emphases, and formulations have, however, presented problems. They have raised questions in the minds of both language teachers and theoretical linguists. How far can language teaching be guided by linguistic theory? To what extent do these changes make more than a superficial contribution to language-teaching practice? Because of the violent controversies in linguistic theory during the 1960s, this problem has been discussed repeatedly between 1966 and the early 1970s. In general, the advice that language teachers have been given is to adopt an independent stance towards linguistic theory—in other words, the resource model. This is, however, much more easily said than done because in order to choose from the many ideas in linguistics and psycholinguistics, the teacher must have a much more wide-ranging and sophisticated knowledge of the language sciences than it is fair to expect.

Another problem can be illustrated by the recent growth of sociolinguistics. This relatively new field epitomized the problem of relating new theory and research to language teaching. Sociolinguistics, as a growing field of study, is transforming our view of language; it is an area that in principle is highly relevant to language teaching. However, theoretical insights outdistance our capabilities to turn them into practical teaching ideas. As so often happens, linguists have developed new concepts and formulated a new model, such as the model of the speech act; but there are, as yet, few data to support it, let alone data for the commonly taught languages. Yet, we are already clamoring for teaching materials embodying these new ideas and modifications in procedures.

Here lurks the danger that language teaching has run into again and again in its relationship to theoretical linguistics (as well as to psychology). There is a shift in theoretical outlook, perfectly legitimate in the development of scientific theory, and to accommodate it—or rather, to keep up with the Joneses—language pedagogy makes slight adjustments in its techniques. A few new terms are sprinkled across the old fare without any fundamental rethinking of the relative importance of the new concepts to the established practice. This is not always so, but at times these changes

may be quite superficial, leaving language teaching not very much the wiser. Such are the dangers. On the positive side, however, linguistic theory challenges our view of language and can in the long run have a profound and beneficial influence on curriculum content and the treatment of language in language teaching. The application model has been used here, sometimes with excellent results, at other times too uncritically.

7.2 *The Impact of Descriptive Linguistics*

The relevance of descriptive linguistics was fully realized and acted upon by linguists of the 1940s. For example, much current teaching of English as a second language is still based on the research done by Fries and Lado in the 1940s and 1950s. Contrastive analysis was an attempt to bring descriptive linguistics even closer to the needs of language teaching. To be sure, descriptive and contrastive linguistics have suffered something of an eclipse as a result of the preoccupation with theory-building between 1965 and 1970, but language teaching cannot do without descriptive investigations if it is to remain up to date, authentic, and reliable. Fortunately, some linguists have not been too much influenced by the changing theoretical emphases and have been prepared to continue a task in which they believe. Otherwise we would not have such recent works of descriptive scholarship as *A Grammar of Contemporary English* (Quirk et al., 1970). Description is particularly necessary today, even in the languages most widely taught, because contemporary views of language are broadening and include an understanding of sociolinguistic and cultural aspects. The fact that this kind of research is in short supply illustrates the problem that priorities in linguistics do not necessarily coincide with those of language teaching.

Several applied linguists around the world have tried to come to grips with the further problem of making language descriptions accessible and useful for language teaching. Some have suggested a data bank of linguistic information so that the course-writer or teacher has a source of reliable and up-to-date linguistic information constantly at his disposal. The data, of course, cannot remain in a data bank alone; hence the idea of a "pedagogical grammar" and of a "sociolinguistic and sociocultural guide," which would serve as a pool of ordered information on language, society, and culture and be the necessary intermediary between theoretical and descriptive research and curriculum development, illustrating the advantages of a convergent model in the relationship between the language sciences and language teaching.

7.3 *The Impact of Psychology and Psycholinguistics*

The influence of psychology and psycholinguistics has taken a different direction from that of linguistics. While in the latter field new concepts

have been formulated without specific regard for second-language teaching and have then been adopted or tried out in language teaching, psychological research has been more responsive to specific problems in education and has sought to confirm and refine notions commonly accepted in the language classroom. Thus the ideas of a "talent for languages" or a "will to learn" were not invented by psychologists but have been defined more precisely in terms of the concepts of aptitude and motivation and can now be related in a more conscious way to other factors in the psychology of language learning and of learning in general. The earlier failure of psychology, and to a certain extent psycholinguistics, to recognize the process of second-language learning as a problem of theoretical interest has, until recently, held up progress in the development of a theory of second-language learning. In this context, it is noteworthy that psychologists did so little for so long to modify the oversimplified view of language learning that has been widespread in language pedagogy. But this deficiency is now being remedied.

Among important contributions we should stress the following: (a) The understanding of second-language skills (comprehension and production) has been deepened by psycholinguistic analyses; (b) the study of first-language acquisition has stimulated empirical investigation of whether it can serve as a model for second-language learning; (c) an important issue now receiving attention is whether there are developmental sequences in first- and second-language learning to which language teaching should be geared; (d) certain learner characteristics—aptitude, motivation, and individual learning styles—have been investigated, and their relationships to each other and to ultimate success in language learning are now better understood; and (e) recent studies of the learning process itself can be expected to lead to important new insights.

On the other hand, some of the most vital issues remain unresolved: (a) The optimal-age issue is still controversial; (b) the widespread failure to learn a second language, although empirically obvious, remains theoretically as puzzling as the inevitability of first-language acquisition; (c) questions about individual differences in language learning and language learner characteristics are still open; and (d) the psychological bases of many second-language teaching techniques are not clear.

Overall, the contact with psychology and psycholinguistics has contributed concepts, models, and basic theoretical positions; it has defined important questions and has suggested possible interpretations. The total volume of psychological and psycholinguistic studies has indirectly created a greater sensitivity to various psychological factors in language learning. At the same time, it clearly shows that the psychological problems of L2 learning demand direct study. They cannot be resolved by extrapolation from other psychological studies.

7.4 *The Impact of Educational Research*

A last major way in which research has made an impact on language teaching is through studies in institutional settings. These studies have demonstrated that research is not necessarily done elsewhere and then "applied" but that it can deal directly with the issues of language teaching in the classroom. In the past, as was pointed out, language teaching questions were not investigated with research methods. Thanks to this group of studies there is wider recognition that a research component may have something to contribute to every phase of the language-teaching operation and may assist in decision making. In all, these studies have yielded a much more realistic understanding of language teaching.¹⁸ The research has made clearer what the options in language teaching are. It has provided a sobering check on the claims—often wild claims—that language-teaching innovators and theorists have been in the habit of making. A far less dogmatic view of teaching methodology has spread. One would be hard put to it today to claim that any narrow set of techniques, specific technological invention, or particular starting age is the "open sesame" of language teaching. In spite of the violent controversies that have surrounded such research, the sobering findings of studies such as the Pennsylvania Project or the Keating Report have largely been confirmed by the test of time, and it is hardly conceivable that major changes in language education could occur in the future without taking account of research.

7.5 *Assessment of Mechanisms*

The language sciences focus on language and language behavior in general; that is their strength. But from the point of view of language pedagogy, it is also a weakness, for it is in the nature of language teaching to be concerned with a particular language. Thus the problem is that those who are concerned with language in general do not themselves teach languages, while those who teach individual languages lack the general theoretical background that would make them more responsive to language-teaching research needs.

It is our contention that in order for language research to make a substantial impact on language teaching, it must eventually capture the interest of those directly concerned with the teaching of particular languages in schools and universities. Retrospectively, it appears to us that the single most important cause of the discrepancy between language-research needs and the systematic provision for research and training is the failure of modern language departments in universities throughout the world to have radically revised their role in language research and

¹⁸ Outstanding examples of the value of research in school settings are the IEA studies of English and French as second languages (Lewis & Massad, 1975; Carroll, 1975b).

language education.

Where, then, did the initiative come from for relevant research done so far? Governments, charitable foundations, university departments of psychology and education, and newly created departments of linguistics have to a large extent displayed the initiative that language departments, with a few exceptions, appear to have lacked. But governments are subject to political pressure; their language interest may rise and fall in unpredictable ways, and research funding, while it has often been unexpectedly generous, may become too erratic to allow for the continuing growth that is needed. There may, for example, be a preference for funding projects that have high "visibility" or that promise to produce "instant" results, rather than studies that fit into a planned framework that takes several years to develop. For psychology or linguistics departments, second-language learning can only be a partial interest if it is not to unbalance seriously their general terms of reference. Under these circumstances, it is surprising how much has so far been achieved in different places around the world in building up a research effort and theoretical discussions of considerable scope.

8. CONCLUSION

We may now inquire what the benefits of this association, of language teaching and research, have been. Has it led to more effective language learning? If we adopt a very critical view of the relations between the research and practice we have reviewed, we find very little concrete evidence that would associate specific improvements in language learning with specific results of research. If we compare the effects of language research with the effects of agricultural, pharmaceutical, or industrial research, we find that the evidence equivalent to an increased crop yield, a complete medical cure, or heightened industrial productivity is lacking in language teaching. The easiest excuse to give would be that the input-output model used in other kinds of applied research is not applicable. It is indeed perhaps unjustified, as Macnamara (1974) has pointed out, to judge the results of language research with the yardstick of commercial productivity.¹⁹ Nevertheless, we want to be assured that there have been true benefits from the research activity, even if they cannot be measured accurately. This case study would seem to indicate that the impact of language-related research has generally been positive, while the dramatic advances in language teaching that some had hoped for have not materialized and some distinct problems have arisen. In making an overall assessment of benefits and defects, we have to bear in mind the cautions

¹⁹ As Macnamara (1974) puts it: "Somehow education in the U.S. has fallen into the hands of businessmen who demand that inputs, processes and outputs be specified in quantified terms" (p. 47).

already expressed in the introduction: It is not possible to provide incontrovertible evidence as proof of particular assertions.

In attempting to assess the merits and demerits of this prolonged contact between the language sciences and language teaching, we cannot proceed as if we were assessing something deliberately planned that could, therefore, be turned on and off at will. Retrospectively, we can examine the evolution of the relationship—as we have done in this case study—and ask ourselves how it came about, whether the influence has been beneficial or not, and whether, now that we are more conscious of the effect of certain influences, we can henceforth deliberately shape the relations between language research and language teaching to enhance their effectiveness.

After events have taken place, one is inclined to take them for granted. Thus, we do not question enough the prolonged interaction between language research and language pedagogy. In some ways, it is surprising that the impact has been as great as it has. What, then, brought language teaching and research together? Research has not come into language teaching as the result of an overwhelming clamor from language teachers. The language teaching profession has, on the whole, long been aware of the difficulties and failures in its endeavors, but in seeking solutions its tendency, until recent decades, has been to speculate, to rely on personal opinion and experience, to try out new methods and institute reforms mainly on a trial-and-error basis, and to resolve ideological conflicts by polemic and debate. Nor did the development of linguistics point inevitably to language teaching as an obvious field of application. Between the 1930s and 1950s, linguistics, as a relatively new science, was more concerned with establishing itself as an autonomous field of study than with applications. However, when in the early 1940s the need arose for practical applications of linguistics to language teaching, linguists were eager to demonstrate the relevance of their discipline. Returning then to the question of what has brought language teaching and research together, we believe it is largely the failure of language teaching to solve its persistent problems in any other way. And it is the equally persistent need, and demand, to come to grips with these problems that has brought about an even closer and more regular association between language research and second-language teaching.

In order to benefit from a science, the educator must regularly, and critically, scan and review the science in question. The scientist who claims to influence the educator with his scientific theories and research findings must try to understand the educational activity in its full complexity, as it occurs. However, the relationship is less productive if the science imposes itself on the educational activity, or if its dictates are uncritically accepted by the educator. Briefly, the conclusion we have reached is that the relationship between research and practice is most fruitful (a) when the

scientist recognizes the full scope of the educational activity in its own right, (b) when the educator, in turn, recognizes the scope of the scientific discipline in its own right, and (c) when the relationship is seen as a dialogue between educator and scientist and not as a monologue by either.

REFERENCES

- Agard, F., & Dunkel, H. B. *An investigation of second language teaching*. Boston: Ginn, 1948.
- Aid, F. M. Semantic universals in instructional materials. *TESOL Quarterly*, 1974, **8**, 53-64.
- Allen, J. P. B. Pedagogic grammar. In J. P. B. Allen & S. P. Corder (Eds.), *Techniques in applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 3). London: Oxford University Press, 1974.
- Allen, J. P. B. Some basic concepts in linguistics. In J. P. B. Allen & S. P. Corder (Eds.), *Papers in applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 2). London: Oxford University Press, 1975.
- Allen, J. P. B., & Corder, S. P. (Eds.). *Readings for applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 1). London: Oxford University Press, 1973.
- Allen, J. P. B., & Corder, S. P. (Eds.). *Techniques in applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 3). London: Oxford University Press, 1974.
- Allen, J. P. B., & Corder, S. P. (Eds.). *Papers in applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 2). London: Oxford University Press, 1975.
- Allen, J. P. B., & Widdowson, H. G. English in physical science. In J. P. B. Allen & H. G. Widdowson (Eds.), *English in focus series*. London: Oxford University Press, 1974.
- Andersson, T. After FLES—What? *Educational Forum*, 1961, **26**, 84-85.
- Anisfeld, M. Psycholinguistic perspectives on language learning. In A. Valdman (Ed.), *Trends in language teaching*. New York: McGraw-Hill, 1966.
- Anisfeld, M., & Lambert, W. E. Social and psychological variables in learning Hebrew. *Journal of Abnormal and Social Psychology*, 1961, **63**, 524-529.
- Asher, J., & Garcia, R. The optimal age to learn a foreign language. *Modern Language Journal*, 1969, **5**, 334-341.
- Austin, J. L. *How to do things with words*. Cambridge, Mass.: Harvard University Press, 1962.
- Bailey, N., Madden, C., & Krashen, S. Is there a "natural sequence" in second language learning? *Language Learning*, 1974, **24**, 235-243.
- Banathy, B., Trager, E. C., & Waddle, C. D. The use of contrastive data in foreign language course development. In A. Valdman (Ed.), *Trends in language teaching*. New York: McGraw-Hill, 1966.
- Barrie, W. B. Semantics and the teaching of vocabulary. *Bulletin CILA*, 1972, **16**, 28-48.
- Birkmaier, E. M. Research on teaching foreign languages. In R. M. Travers (Ed.), *Second handbook of research on teaching: A project of the American Education Research Association*. Chicago: Rand-McNally, 1973.
- Bloch, B., & Trager, G. L. *Outline of linguistic analysis*. Baltimore: Linguistic Society of America, 1942.
- Bloomfield, L. *Language*. New York: Holt, Rinehart & Winston, 1933.

- Bloomfield, L. *Outline guide for the practical study of foreign languages*. Baltimore: Linguistic Society of America, 1942.
- Bolinger, D. L. The atomization of meaning. *Language*, 1965, **37**, 366-381.
- Bolinger, D. L. The theorist and the language teacher. *Foreign Language Annals*, 1968, **2**, 30-41.
- Bolinger, D. L. The influence of linguistics: Plus and minus. *TESOL Quarterly*, 1972, **6**, 107-120.
- Bolinger, D. L., Ciruti, J. E., & Montero, H. H. *Modern Spanish: A project of the Modern Language Association*. New York: Harcourt, Brace, 1960.
- Brooks, N. *Language and language learning—Theory and practice*. New York: Harcourt, Brace, 1960.
- Brown, H. D. The psychological reality of "grammar" in the ESL classroom. Paper presented at the Sixth Annual TESOL Convention, Washington, D.C., February 29, 1972. (ED 061 845)
- Brown, R. *A first language: The early stages*. Cambridge, Mass.: Harvard University Press, 1973.
- Brown, T. G. Pedagogical implications of a case grammar of French. *IRAL*, 1971, **9**, 229-244.
- Buchanan, M. A., & MacPhee, E. D. *An annotated bibliography of modern language methodology*. Toronto: University of Toronto Press, 1928.
- Burstall, C. *French from eight: A national experiment*. Slough, England: National Foundation for Educational Research, 1968.
- Burstall, C. *French in the primary school: Attitudes and achievement*. Slough, England: National Foundation for Educational Research, 1970.
- Burstall, C. Factors affecting foreign-language learning: A consideration of some recent research findings. *Language Teaching and Linguistic Abstracts*, 1975, **8**, 5-25.
- Burstall, C., Jamieson, M., Cohen, S., & Hargreaves, M. *Primary French in the balance*. Slough, England: National Foundation for Educational Research, 1974.
- Carroll, J. B. *The study of language: A survey of linguistics and related disciplines in America*. Cambridge, Mass.: Harvard University Press, 1953.
- Carroll, J. B. Wanted: A research basis for educational policy on foreign language teaching. *Harvard Educational Review*, 1960, **30**, 128-140.
- Carroll, J. B. The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education*. Pittsburgh: University of Pittsburgh Press, 1962.
- Carroll, J. B. A model of school learning. *Teachers College Record*, 1963, **64**, 723-733. (a)
- Carroll, J. B. Research on teaching foreign languages. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand-McNally, 1963. (b)
- Carroll, J. B. The contributions of psychological theory and educational research to the teaching of foreign languages. In A. Valdman (Ed.), *Trends in language teaching*. New York: McGraw-Hill, 1966.
- Carroll, J. B. Research problems concerning the teaching of foreign or second languages to younger children. In H. H. Stern (Ed.), *Foreign languages in primary education*. London: Oxford University Press, 1967.
- Carroll, J. B. Guide for the collection of data pertaining to the study of foreign or second languages by younger children. In H. H. Stern (Ed.), *Languages and the young school child*. London: Oxford University Press, 1969. (a)
- Carroll, J. B. Psychological and educational research into second language teaching to younger children. In H. H. Stern (Ed.), *Languages and the young school child*. London: Oxford University Press, 1969. (b)

- Carroll, J. B. Current issues in psycholinguistics and second-language teaching. *TESOL Quarterly*, 1971, **5**, 101-114.
- Carroll, J. B. Learning theory for the classroom teacher. In G. A. Jarvis (Ed.), *The challenge of communication. The ACTFL review of foreign language education* (Vol. 6). Skokie, Ill.: National Textbook, 1974.
- Carroll, J. B. Aptitude in second language learning. In G. Taggart (Ed.), *Attitude and aptitude in second language learning*. Canadian Association of Applied Linguistics, Fifth Symposium, May 1974; *Proceedings*, 1975. (a)
- Carroll, J. B. *The teaching of French as a foreign language in eight countries*. New York: Wiley, 1975. (b)
- Carroll, J. B., & Sapon, S. M. *Modern language aptitude test* (MLAT). New York: Psychological Corporation, 1959.
- Carroll, J. B., & Sapon, S. M. *Modern language aptitude test—Elementary* (EMLST). New York: Psychological Corporation, 1967.
- Carton, A. S. Inferencing: A process in using and learning language. In P. Pimsleur & T. Quinn (Eds.), *The psychology of second language learning*. Cambridge, England: Cambridge University Press, 1971.
- Chafe, W. *Meaning and the structure of language*. Chicago: University of Chicago Press, 1970.
- Chastain, K. *The development of modern language skills: Theory to practice*. Philadelphia: Center for Curriculum Development, 1971.
- Chastain, K. D., & Woerdehoff, F. J. A methodological study comparing the audio-lingual habit theory and the cognitive code-learning theory. *Modern Language Journal*, 1968, **52**, 268-279.
- Chomsky, N. *Syntactic structures*. The Hague: Mouton, 1957.
- Chomsky, N. Review of B. F. Skinner, *Verbal Behavior*. *Language*, 1959, **35**, 26-53.
- Chomsky, N. *Aspects of the theory of syntax*. Cambridge, Mass.: M.I.T. Press, 1965.
- Chomsky, N. Linguistic theory. In R. G. Mead, Jr. (Ed.), *Language teaching: Broader contexts*. New York: Northeast Conference on the Teaching of Foreign Languages, 1966.
- Clifford, J. C. A history of the impact of research on teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand-McNally College Publishing, 1973.
- Connolly, P. C. How to teach families of words by comparison. *English Language Teaching*, 1973, **27**, 171-176.
- Cooke, D. *The role of explanation in foreign language instruction*. Unpublished doctoral dissertation, University of Essex, 1974.
- Corder, S. P. *Introducing applied linguistics*. Harmondsworth, England: Penguin, 1973.
- Corder, S. P. The significance of learner's errors. In J. C. Richards (Ed.), *Error analysis: Perspectives on second language acquisition*. London: Longman, 1974. (Originally published, 1967; *IRAL*, 1967, **5**, 161-170.)
- Corder, S. P. *The study of interlanguage*. Paper delivered at the Fourth AILA Congress, Stuttgart, 1975.
- Cowan, S. English proficiency and bicultural attitudes of Japanese students. *English Teacher's Magazine*, 1968, **17**, 38-44.
- CREDIF. *Voix et images de France*. Philadelphia: Chilton, 1962.
- Criper, C., & Widdowson, H. G. Sociolinguistics and language teaching. In J. P. B. Allen & S. P. Corder (Eds.), *Papers in applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 2). London: Oxford University Press, 1975.
- Crothers, E., & Suppes, P. *Experiments in second-language learning*. New York: Academic Press, 1967.

- Curran, C. A. *Counseling-learning: A whole-person model for education*. New York: Grune & Stratton, 1972.
- de Villiers, J., & de Villiers, P. A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, 1973, **2**, 267-278.
- Diller, K. C. Some new trends for applied linguistics and foreign language teaching in the United States. *TESOL Quarterly*, 1975, **9**, 66-73.
- Di Pietro, R. *Language structures in contrast*. Rowley, Mass.: Newbury House, 1971.
- Disick, R. S., & Barbanel, L. Affective education and foreign language learning. In G. A. Jarvis (Ed.), *The challenge of communication: The ACTFL review of foreign language education* (Vol. 6). Skokie, Ill.: National Textbook, 1974.
- Dulay, H. C., & Burt, M. K. Should we teach children syntax? *Language Learning*, 1973, **23**, 245-258.
- Dulay, H. C., & Burt, M. K. Errors and strategies in child second language acquisition. *TESOL Quarterly*, 1974, **8**, 129-136. (a)
- Dulay, H. C., & Burt, M. K. Natural sequences in child second language acquisition. *Language Learning*, 1974, **24**, 37-53. (b)
- Dulay, H. C., & Burt, M. K. A new perspective on the creative construction process in child second language acquisition. *Working Papers on Bilingualism*, 1974, **4**, 71-98. (c)
- Dulay, H. C., & Kiparsky, C. *The gooficon: A repair manual for English*. Rowley, Mass.: Newbury House, 1972.
- Ervin-Tripp, S. M. An Issei learns English. *Journal of Social Issues*, 1967, **23**, 78-90.
- Ervin-Tripp, S. M. Is second language learning like the first? *TESOL Quarterly*, 1974, **8**, 111-127.
- Feenstra, H. J. *Aptitude, attitude and motivation in second-language acquisition*. Unpublished doctoral dissertation, University of Western Ontario, 1967.
- Feenstra, H. J., & Gardner, R. C. *Aptitude, attitude and motivation in second-language acquisition* (Research Bulletin 101). London, Canada: University of Western Ontario, 1968.
- Fillmore, C. The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart & Winston, 1968.
- Fishman, J. A. The implications of bilingualism for language teaching and language learning. In A. Valdman (Ed.), *Trends in language teaching*. New York: McGraw-Hill, 1966.
- Fishman, J. A. *Sociolinguistics: A brief introduction*. Rowley, Mass.: Newbury House, 1970.
- Fishman, J. A. *The sociology of language: An interdisciplinary social science approach to language in society*. Rowley, Mass.: Newbury House, 1972.
- Fishman, J. A., Cooper, R., & Ma, R. *Bilingualism in the barrio* (Language Science Monographs). Bloomington: Indiana University, 1971.
- Ford Foundation. *Language and development: A retrospective survey of Ford Foundation language projects 1952-1974. Case studies*. New York: Ford Foundation, 1975.
- Fox, M. J. *Language education in developing countries: The changing role of the Ford Foundation*. New York: Ford Foundation, 1974.
- Fox, M. J. *Language and development: A retrospective survey of Ford Foundation language projects, 1952-1974*. New York: Ford Foundation, 1975.
- Fox, M. J., & Skolnick, B. P. *Language in education: Problems and prospects in research and training*. New York: Ford Foundation, 1975.
- Fries, C. C. *Teaching and learning English as a foreign language*. Ann Arbor: University of Michigan Press, 1945.

- Fries, C. C. The Chicago investigation. *Language Learning*, 1949, **2**, 89-99.
- Fries, C. C. *The structure of English: An introduction to the construction of English sentences*. New York: Harcourt, Brace, 1952.
- Gardner, R. C. *Motivational variables in second-language acquisition*. Unpublished doctoral dissertation, McGill University, 1960.
- Gardner, R. C. *Motivation and second-language acquisition* (Research Bulletin 276). London, Canada: University of Western Ontario, Department of Psychology, 1973.
- Gardner, R. C., & Lambert, W. E. Motivational variables in second-language acquisition. *Canadian Journal of Psychology*, 1959, **13**, 266-272.
- Gardner, R. C., & Lambert, W. E. *Attitudes and motivation in second-language learning*. Rowley, Mass.: Newbury House, 1972.
- Gardner, R. C., & Santos, E. H. *Motivational variables in second-language acquisition: A Philippine investigation* (Research Bulletin 149). London, Canada: University of Western Ontario, 1970.
- Gardner, R. C., Taylor, D. M., & Feenstra, H. J. Ethnic stereotypes: Attitudes and beliefs. *Canadian Journal of Psychology*, 1970, **24**, 32-34.
- Geseli, A., & Ilg, F. L. Development trends in language behavior. *FL Bulletin*, 1956, **49**, 6-8. (Reissued, 1961.)
- Gimson, A. C. *An introduction to the pronunciation of English*. London: Arnold, 1970.
- Gleason, H. A., Jr. Contrastive analysis in discourse structure. In J. E. Alatis (Ed.), *Report of the Nineteenth Annual Round Table Meeting on Linguistics and Language Study*. Washington, D.C.: Georgetown University, 1968.
- Gorbet, F. Error analysis: What the teachers can do. In *Errors: A new perspective*. Ottawa: Bureau of Staff Development and Training, Language Bureau, Research Division, 1974.
- Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageat, A. *L'Elaboration du Français fondamental* (1^{er} degré). Paris: Didier, 1964.
- Gouin, F. *The art of teaching and studying languages*. New York: Scribner, 1892.
- Grauberg, W. An error analysis in German of first-year university students. In G. E. Perrier & J. L. M. Trim (Eds.), *Applications of linguistics*. London: Cambridge University Press, 1971.
- Grittner, F. M. *Teaching foreign languages*. New York: Harper & Row, 1969.
- Hakuta, K. A preliminary report on the development of grammatical morphemes in a Japanese girl learning English as a second language. *Working Papers on Bilingualism*, 1974, **3**, 18-43.
- Halliday, M. A. K. Notes on transitivity and theme in English. *Journal of Linguistics*, 1967, **3**, 37-81, 199-244; 1968, **4**, 179-215. (A three-part article.)
- Halliday, M. A. K. Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics*. Harmondsworth, England: Penguin, 1970.
- Halliday, M. A. K., McIntosh, A., & Stevens, P. *The linguistic sciences and language teaching*. London: Longmans, 1964.
- Hancock, C. R. Student aptitude: Predictive value. In G. Taggart (Ed.), *Attitude and aptitude in second language learning*. Toronto: Canadian Association of Applied Linguistics, 1975.
- Hansen-Bede, L. A child's creation of a second language. *Working Papers on Bilingualism*, 1975, **6**, 103-126.
- Hatch, E. Second language learning universals? *Working Papers on Bilingualism*, 1974, **3**, 1-17.
- Hendricks, W. O. Current trends in discourse analysis. In B. Kachru & H. F. W. Stahlke, *Current trends in stylistics*. Champaign, Ill.: Linguistic Research, 1972.

- Hubbard, L. J. Aptitude, attitude and sensitivity. *Foreign Language Annals*, 1975, 8, 33-37.
- Hymes, D. *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press, 1974.
- Jakobovits, L. A. *Foreign language learning: A psycholinguistic analysis of the issues*. Rowley, Mass.: Newbury House, 1970.
- Jakobovits, L. A., & Gordon, B. *The context of foreign language teaching*. Rowley, Mass.: Newbury House, 1974.
- Jarvis, R. A. A pedagogical grammar of the modal auxiliaries. *English Language Teaching*, 1972, 26, 238-248.
- Jonas, S. R. *A matched classroom approach to the teaching of French in the elementary grades*. Cincinnati: College of Mount St. Joseph on the Ohio, 1971.
- Jones, G. F. (Ed.). *Foreign language teaching: Ideals and practices*. New York: Northeast Conference on the Teaching of Foreign Languages, 1964.
- Katz, J., & Fodor, J. The structure of a semantic theory. *Language*, 1963, 39, 170-210.
- Katz, J., & Postal, P. *An integrated theory of linguistic descriptions*. Cambridge: M.I.T. Press, 1964.
- Keating, R. F. *A study of the effectiveness of language laboratories*. New York: Columbia University Teachers College, Institute of Administrative Research, 1963.
- Kellerman, E. Elicitation, lateralization and error analysis. In *York papers in linguistics*. York, England: York University, 1974.
- Kelly, L. G. *Twenty-five centuries of language teaching*. Rowley, Mass.: Newbury House, 1969.
- Khampang, P. Thai difficulties in using English prepositions. *Language Learning*, 1974, 24, 215-222.
- Lado, R. *Language testing*. London: Longmans, 1961.
- Lado, R. *Language teaching: A scientific approach*. New York: McGraw-Hill, 1964.
- La Forge, P. G. Community language learning: A pilot study. *Language Learning*, 1971, 21, 45-61.
- Lambert, W. E. Psychological studies of interdependencies of the bilingual's two languages. In J. Puhvel (Ed.), *Substance and structure of language*. Los Angeles: University of California Press, 1969.
- Lambert, W. E., Gardner, R. C., Barik, H. C., & Tunstall, K. Attitudinal and cognitive aspects of intensive study of a second language. *Journal of Abnormal and Social Psychology*, 1963, 66, 358-369.
- Lambert, W. E., & Tucker, G. R. *Bilingual education of children: The St. Lambert experiment*. Rowley, Mass.: Newbury House, 1972.
- Lamendella, J. T. On the irrelevance of transformational grammar to second language pedagogy. *Language Learning*, 1969, 19, 255-270.
- Léon, M. *Exercices systématiques de prononciation Française*. Paris: Librairies Hachette et Larousse, 1964.
- Léon, P. R. *Prononciation du Français standard*. Paris: Didier, 1966.
- Levenston, E. A. Aspects of testing the oral proficiency of adult immigrants to Canada. *TESL Talk*, 1973, 4, 39-48. (a)
- Levenston, E. A. *Teaching indirect object structures in English—A case study in applied linguistics*. Unpublished manuscript, Modern Language Centre, OISE, Toronto, 1973. (b)
- Levin, L. *Comparative studies in foreign-language teaching: The GUME project*. Stockholm: Almqvist & Wiksell, 1972.
- Lewis, E. G. *Linguistics and second language pedagogy: A theoretical study*. The Hague: Mouton, 1974.

- Lewis, E. G., & Massad, C. E. *English as a foreign language in ten countries*. New York: Wiley, 1975.
- Lyons, J. *Structural semantics*. Oxford: Blackwell, 1963.
- Lyons, J. *Introduction to theoretical linguistics*. London: Cambridge University Press, 1968.
- Mackey, W. F. *Language teaching analysis*. London: Longmans, 1965.
- Mackey, W. F. *Bilingualism as a world problem*. Montreal: Harvest House, 1967.
- Macnamara, J. The cognitive basis of language learning in infants. *Psychological Review*, 1972, 79, 1-13.
- Macnamara, J. What can we expect from a bilingual program? *Working Papers on Bilingualism*, 1974, 4, 42-56.
- McCawley, J. D. The role of semantics in a grammar. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart & Winston, 1968.
- McKinnon, K. R. *An experimental study of the learning of syntax in second language learning*. Unpublished doctoral dissertation, Harvard University, 1965.
- McNeill, D. The creation of language by children. In J. Lyons & R. M. Wales (Eds.), *Psycholinguistic papers*. Edinburgh: Edinburgh University Press, 1966.
- Mougeon, R., & Hébrard, P. *L'Acquisition et la maîtrise de l'anglais parlé par les jeunes bilingues de Welland*. Unpublished report, Section Franco-Ontarienne, OISE, Toronto, 1975.
- Moulton, W. G. Linguistics and language teaching in the United States 1940-1960. In C. Mohrmann, A. Sommerfelt, & J. Whatmough (Eds.), *Trends in European and American linguistics*. Utrecht: Spectrum, 1961.
- Moulton, W. G. *The sound of English and German*. Chicago: University of Chicago Press, 1962.
- Naiman, N., Fröhlich, M., & Stern, H. H. *The good language learner*. Unpublished report to the Secretary of State, 1975.
- Naiman, N., Fröhlich, M., & Todesco, A. The good language learner. *TESL Talk*, 1975, 6, 58-76.
- Nelson, R. J. Motivation in foreign-language learning. In J. A. Turst (Ed.), *Foreign languages and the "new" student*. New York: Northeast Conference on the Teaching of Foreign Languages, 1970.
- Nemser, W. Approximative systems of foreign language learners. *IRAL*, 1971, 9, 115-123.
- Neufeld, G. G. Foreign language aptitude: An enduring process. In G. Rondeau (Ed.), *Some aspects of Canadian applied linguistics*. Montreal: Centre Educatif et Culturel, 1973.
- Neufeld, G. G. A theoretical perspective on the nature of linguistic aptitude. In G. Taggart (Ed.), *Attitude and aptitude in second language learning*. Toronto: Canadian Association of Applied Linguistics, 1975.
- Newmark, L. How not to interfere with language learning. In M. Lester (Ed.), *Readings in applied transformational grammar*. New York: Holt, Rinehart & Winston, 1970. (Reprinted from the *International Journal of American Linguistics*, 1966, 32, 77-83.)
- Newmark, L. A minimal language teaching program. In P. Pimsleur & T. Quinn (Eds.), *The psychology of second language learning*. London: Cambridge University Press, 1971.
- Newmark, L., & Reibel, D. Necessity and sufficiency in language learning. *IRAL*, 1968, 6, 145-161.
- Nickel, G. (Ed.). *Papers in contrastive linguistics*. London: Cambridge University Press, 1971.
- Nida, E. A. *Learning a foreign language*. Ann Arbor: Friendship Press, 1957.

- Nilsen, D. L. F. The use of case grammar in teaching English as a foreign language. *TESOL Quarterly*, 1971, **5**, 293-299.
- Noblitt, J. S. Pedagogical grammar: Towards a theory of foreign language materials preparation. *IRAL*, 1972, **10**, 313-329.
- Norris, W. E. *TESOL at the beginning of the '70's: Trends, topics, and research needs*. Pittsburgh: University of Pittsburgh, 1971.
- Oller, J. *Second language testing*. Seminar and paper presented at the Sixth Annual Conference on Applied Linguistics, University of Michigan, January 31 to February 1, 1975.
- Olson, D. R. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 1970, **77**, 257-273.
- Osgood, C. E., & Sebeok, T. A. (Eds.). *Psycholinguistics: A survey of theory and research problems*. Baltimore: Waverly Press, 1954.
- Palmer, H. E. *The principles of language study*. London: Oxford University Press, 1964. (Originally published, 1922.)
- Palmer, H. E. *The scientific study and teaching of languages*. London: Oxford University Press, 1968. (Originally published, 1917.)
- Parker, W. R. *The national interest and foreign languages*. Washington, D.C.: U.S. Government Printing Office, 1961.
- Paulston, C. B. A biased bibliography: Comments on selecting texts for a methods course in TESOL. *Language Learning*, 1973, **23**, 129-143.
- Paulston, C. B. Linguistic and communicative competence. *TESOL Quarterly*, 1974, **8**, 347-362.
- Paulston, C. B. Ethnic relations and bilingual education: Accounting for contradictory data. *Working Papers on Bilingualism*, 1975, **6**, 1-44.
- Penfield, W. G. A consideration of neurophysiological mechanisms of speech and some educational consequences. In *Proceedings of the American Academy of Arts and Sciences*, 1953, **82**, 199-214.
- Penfield, W. G. Conditioning the uncommitted cortex for language learning. *Brain*, 1965, **88**, 787-798.
- Penfield, W. G., & Roberts, L. *Speech and brain mechanisms*. Princeton, N.J.: Princeton University Press, 1959.
- Pillet, R. A. *Foreign-language study: Perspective and prospect*. Chicago: University of Chicago Press, 1974.
- Pimsleur, P. (Coordinator). *Report of the conference on psychological experiments related to second-language learning* (December 29, 1959 to January 1, 1960). Los Angeles: University of California, 1960.
- Pimsleur, P. *Pimsleur language aptitude battery*. New York: Harcourt, Brace, Jovanovitch, 1966.
- Pimsleur, P., Stockwell, R. P., & Comrey, A. L. Foreign language learning ability. *Journal of Educational Psychology*, 1962, **53**, 15-26.
- Pimsleur, P., Sundland, D. M., & McIntyre, R. D. *Underachievement in foreign language learning*. New York: MLA Materials Center, 1966.
- Politzer, R. L. *Teaching French: An introduction to applied linguistics*. New York: Ginn, 1960.
- Politzer, R. L. *Foreign language learning: A linguistic introduction*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- Politzer, R. L. *Linguistics and applied linguistics: Aims and methods*. Philadelphia: Center for Curriculum Development, 1972.
- Prator, C. H., Jr. *A manual of American English pronunciation*. New York: Holt, Rinehart, & Winston, 1957.
- Quinn, T. J. Theoretical foundations in linguistics and related fields. *The ACTFL Review of Foreign Language Education*, 1974, **5**, 329-353.

- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. *A grammar of contemporary English*. London: Longmans, 1970.
- Rand, E. *Constructing dialogs*. New York: Holt, Rinehart & Winston, 1969.
- Ravem, R. The development of wh-questions in first and second language learners. In J. C. Richards (Ed.), *Error analysis: Perspectives on second language acquisition*. London: Longman, 1974.
- Report of the Royal Commission on Bilingualism and Biculturalism*. Ottawa: Queen's Printer, 1967-1970.
- Richards, J. C. (Ed.). *Error analysis: Perspectives on second language acquisition*. London: Longman, 1974.
- Rivers, W. M. *The psychologist and the foreign language teacher*. Chicago: University of Chicago Press, 1964.
- Rivers, W. M. *Teaching foreign language skills*. Chicago: University of Chicago Press, 1968.
- Rivers, W. M. Linguistic and psychological factors in speech perception and their implications for teaching materials. In P. Pimsleur & T. Quinn (Eds.), *The psychology of second language learning*. London: Cambridge University Press, 1971.
- Rivers, W. M. *Speaking in many tongues: Essays on foreign-language teaching*. Rowley, Mass.: Newbury House, 1972.
- Rivers, W. M. *A practical guide to the teaching of French*. London: Oxford University Press, 1975. (a)
- Rivers, W. M. Students, teachers, and the future. *Foreign Language Annals*, 1975, 8, 22-32. (b)
- Rivers, W. M., Dell'orto, K. M., & Dell'orto, V. J. *A practical guide to the teaching of German*. London: Oxford University Press, 1975.
- Roberts, D. A., & Connelly, F. M. The "conceptualization/practice interface": A potentially useful notion for exploring the relevance of graduate study in curriculum. Paper presented at the Annual Meeting of the Canadian Association for Curriculum Studies, Edmonton, June 4, 1975.
- Rosenbaum, P. S. *The grammar of English predicate complement constructions*. Cambridge, Mass.: M.I.T. Press, 1967.
- Roulet, E. *Théories grammaticales, descriptions et enseignement des langues*. Paris: Nathan, 1972.
- Rubin, J. *National bilingualism in Paraguay*. The Hague: Mouton, 1968.
- Rubin, J. What the "good language learner" can teach us. *TESOL Quarterly*, 1975, 9, 41-51.
- Rutherford, W. E. *Modern English: A textbook for foreign students*. New York: Harcourt, Brace & World, 1968.
- Saporta, S. Applied linguistics and generative grammar. In A. Valdman (Ed.), *Trends in language teaching*. New York: McGraw-Hill, 1966.
- Savignon, S. J. *Communicative competence: An experiment in foreign language teaching*. Philadelphia: Center for Curriculum Development, 1972.
- Schachter, J. An error in error analysis. *Language Learning*, 1974, 24, 205-214.
- Scherer, G. A. C., & Wertheimer, M. A. *A psycholinguistic experiment in foreign language teaching*. New York: McGraw-Hill, 1964.
- Schegloff, E. A., & Sacks, H. Opening up closings. *Semiotica*, 1973, 8, 289-327.
- Schumann, J. H. *Second language acquisition research: Getting a more global look at the learner*. Paper presented at the University of Michigan Sixth Annual Conference on Applied Linguistics, Ann Arbor, Michigan, January 31 to February 1, 1975.
- Schwab, J. J. *The practical: A language for curriculum*. Washington, D.C.: National Education Association, 1970.

- Scovel, T. Foreign accents, language acquisition and cerebral dominance. *Language Learning*, 1969, **19**, 245-253.
- Searle, J. *Speech acts: An essay in the philosophy of language*. London: Cambridge University Press, 1969.
- Selinker, L. The application of linguistics to foreign language teaching: The state of the art. In *Reports and papers from the State Conference in Spokane*, March 17-18, 1972. (ED 065015)
- Selinker, L. Interlanguage. In J. C. Richards (Ed.), *Error analysis: Perspectives on second language acquisition*. London: Longman, 1974. (Originally published, 1972.)
- Shuy, R. W. The relevance of sociolinguistics for language teaching. *TESOL Quarterly*, 1969, **3**, 13-22.
- Simmins, C. A. The mental processes involved in learning a foreign language. *Modern Languages*, 1930, **12**, 37-43.
- Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
- Smith, P. D., Jr. *A comparison of the cognitive and audiolingual approaches to foreign language instruction*. Philadelphia: Center for Curriculum Development, 1970.
- Spolsky, B. Attitudinal aspects of second language learning. *Language Learning*, 1969, **19**, 271-285. (a)
- Spolsky, B. Linguistics and pedagogy—Applications or implications? In J. E. Alatis (Ed.), *Report of the Twentieth Annual Round Table Meeting on Linguistics and Language Study*. Washington, D.C.: Georgetown University, 1969. (b)
- Spolsky, B. Linguists and language testers. In B. Spolsky (Ed.), *Current trends in language testing*. The Hague: Mouton, in press.
- Stack, E. M. Advances in language teaching in the United States. In B. Libbush (Ed.), *Advances in the teaching of modern languages* (Vol. 1). Oxford: Pergamon Press, 1964.
- Stern, H. H. *Foreign languages in primary education*. London: Oxford University Press, 1967.
- Stern, H. H. (Ed.). *Languages and the young school child*. London: Oxford University Press, 1969.
- Stern, H. H. *Perspectives on second language teaching*. Toronto: Ontario Institute for Studies in Education, 1970.
- Stern, H. H. Directions in language teaching theory and research. In *Applied linguistics: Problems and solutions, AILA proceedings, 1972*. Heidelberg: Groos, 1974.
- Stern, H. H. What can we learn from the good language learner? *Canadian Modern Language Review*, 1975, **31**, 304-318.
- Stern, H. H., Swain, M., McLean, L. D., Friedman, R. J., Harley, B., & Lapkin, S. *Three approaches to teaching French: Evaluation and overview of studies related to the federally-funded extensions of the second-language learning (French) programs in the Carleton and Ottawa school boards*. Toronto: Ministry of Education, Ontario, 1976.
- Stevick, E. W. Review of Curran 1972. *Language Learning*, 1973, **23**, 259-271.
- Stockwell, R. P., & Bowen, J. D. *The sounds of English and Spanish*. Chicago: University of Chicago Press, 1967.
- Swain, M. French immersion programs across Canada. *Canadian Modern Language Review*, 1974, **31**, 117-129.
- Swain, M., & Barik, H. C. *Five years of primary French immersion: Annual reports of the bilingual education project to the Carleton Board of Education and the Ottawa Board of Education, 1972-1975*. Toronto: Ontario Institute for Studies in Education, 1976.

- Sweet, H. *The practical study of languages*. London: Oxford University Press, 1964. (Originally published, 1899.)
- Symonds, P. M. *Foreign language prognosis test* (Form A & B). New York: Columbia University, Teachers College, 1930.
- Thorndike, E. L. *The teacher's word book*. New York: Columbia University, Teachers College, 1921.
- Thorndike, E. L. *Adult learning*. New York: Macmillan, 1928.
- Thorndike, E. L. On learning of rules in the study of a foreign language. *German Quarterly*, 1931, 4, 89-95.
- Trim, J. L. M. *The analysis of "language contents" for a European unit/credit system in modern languages within the framework of continuing post-secondary adult education*. Strasbourg: Council of Europe, Committee for Out-of-School Education and Cultural Development, 1971.
- Valdman, A. (Ed.). *Trends in language teaching*. New York: McGraw-Hill, 1966.
- van Buren, P. Contrastive analysis. In J. P. B. Allen & S. P. Corder (Eds.), *Techniques in applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 3). London: Oxford University Press, 1974.
- van Buren, P. Semantics and language teaching. In J. P. B. Allen & S. P. Corder (Eds.), *Papers in applied linguistics* (Edinburgh Course in Applied Linguistics, Vol. 2). London: Oxford University Press, 1975.
- van Ek, J. A. *The threshold level in a European unit/credit system for modern language learning by adults*. Strasbourg: Council of Europe, 1975.
- Varadi, T. *Strategies of target language learner communication: Message-adjustment*. Paper presented at the Sixth Conference of the Romanian-English Linguistics Project, Timisoara, Romania, May 1973.
- Wardhaugh, R. *Topics in applied linguistics*. Rowley, Mass.: Newbury House, 1974.
- Wesche, M. B. *The good adult language learner: A study of learning strategies and personality factors in an intensive course*. Unpublished doctoral dissertation, University of Toronto, 1975.
- Widdowson, H. G. Directions in the teaching of discourse. In S. P. Corder & E. Roulet (Eds.), *Theoretical models in applied linguistics*. Paris: Didier, 1973.
- Wilkins, D. A. *Linguistics in language teaching*. Cambridge, Mass.: MIT Press, 1972.
- Wilkins, D. A. *Grammatical, situational and notional syllabuses* (E.L.T. Documents, 73/6). London: English Teaching Information Centre, 1973.
- Yalden, J. Information resources in second-language teaching and learning: A guide for teachers, researchers and educators. *Canadian Modern Language Review*, 1976, 32, 318-348.

Chapter 7

PARADIGM AND PRACTICE: ON THE IMPACT OF BASIC RESEARCH IN EDUCATION

by

J. W. GETZELS

The University of Chicago

THE skepticism of educators about the bearing of research—and particularly of basic or theory-oriented research—on the operation of schools is quite extraordinary. Although the first efforts by the federal government to establish a research component for education came with the passage of the Cooperative Research Act in 1954, the National Education Association (NEA) and its subsidiary, the American Association of School Administrators (AASA), refused to support appropriations for educational research until the passage of the Elementary and Secondary Education Act in 1965, which they were finally persuaded to support by a personal appeal from President Johnson (Stiles, 1972).¹

Even more extraordinary is the skepticism of the researchers themselves about the bearing of what they are doing on the operation of schools. Here, as one instance for many, is a not atypical statement emanating from a prominent educational research center:

The fact is basic research produces only tiny "bits" of knowledge about an infinitely complex world. Given time, such incremental gains in knowledge

¹ Parts of this chapter are based on "Paradigm and Practice: On the Contributions of Research to Education," a paper presented at the American Educational Research Association annual meeting, Los Angeles, California, Feb. 5-8, 1969; "The Impact of Basic Research on Educational Practice," a paper presented at the American Association for the Advancement of Science annual meeting, Boston, Massachusetts, Dec. 30, 1969; "Problem Finding," the 343rd Convocation Address, the University of Chicago (*The University of Chicago Record*, 1973, 7(9), 281-283); "Images of the Classroom and Visions of the Learner" (*School Review*, 1974, 82, 527-540); and "Educational Administration Twenty Years Later: 1954-1974," the keynote address at the Conference of the University Council for Educational Administration, Columbus, Ohio, April 27-30, 1975. I am grateful to Ms. Alison Oyer for her bibliographic assistance and critical reading of the manuscript.

may have some effect on practical situations, but schoolmen have long been aware that research findings usually have little immediate, practical bearing on the actual operation of schools. (*R & D Perspectives*, 1969, p. 1; see also Start, 1975).

On grounds of assertions such as this—that research findings have little practical bearing on the actual operation of schools—support for research in education by foundation and government agencies has been sacrificed for other more “applied” activities, which presumably have instant practical bearing. Take, as one instance only, the federal government. In comparison to support for research and development in other areas, education receives only token funding. Of the almost \$17 billion appropriated for research and development by Congress in 1972, over \$8.3 billion went to defense, \$3.2 billion to space research, \$1.2 billion to atomic energy, but only \$122 million to education (Stiles, 1972). To be sure, some of the expenditure in defense, space, and atomic energy probably went into educational research. And expenditure for educational research, including perhaps even basic work, may also be made at the local level, as it would not be in the other areas. Nonetheless, the dearth of federal funding for education in its own right is noteworthy.

Or looked at another way, in 1971 the nation maintained a \$51.5 billion health enterprise and spent \$2.5 billion on health research and development, almost 4.7% of the total budget, in agriculture the expenditures were \$73.5 billion with \$1.1 billion going to research, over 1.5%. In contrast, education devoted—in all categories—only \$210 million of a \$53 billion budget to research, or less than .4% (Stiles, 1972). Moreover, when research funds are curtailed, as they were for the National Institute of Education (NIE) in 1974, among the first and surest victims is the Field Initiated Studies Program, which includes the preponderance of basic or theory-oriented research (Stivers, 1973). The Field Initiated Studies Program is currently spending less money on the research of individual inquirers than the Cooperative Research Program spent in 1962 (Guba & Clark, 1975).

As Arthur Wise (1976), associate director of the National Institute of Education from 1972 to 1975, pointed out recently, although NIE was created specifically to “foster careful research, the practical application of which would take time,” there was almost from the beginning a “shift in emphasis from research to development, demonstration, and dissemination,” which would presumably promote improvement in the short run. In Wise’s words, “‘Scientific inquiry into educational process’ was the original congressional mandate for the NIE. Now only a small portion of the NIE budget is devoted to serious scientific inquiry” (p. 65). Indeed, “throughout, the proportion of the budget actually expended on research, by generous assumptions, never exceeded 15%” (p. 64).

A brief look at the changes in the internal organizational structure of

NIE reveals how the institute shifted the image of its mission away from research. The first structure consisted of four subunits, all of which contained the word "research" in their titles: Office of Research Grants, Office of Research and Exploratory Studies, Office of Programmatic Research and Development, and Office of Research and Development Resources. The second structure contained "research" in the title of only one subunit. The third and present structure does not contain the word "research" in the title of any subunit: Basic Skills, Education and Work, Finance and Productivity, Educational Equity, School Capacity for Problem Solving, and Dissemination and Resources. The changes in title are not merely nominal. As Wise says, "The choice of organizational nomenclature in the federal bureaucracy is terribly significant, for it is the major means by which the organization is presented; nomenclature reveals organizational priorities and is the primary basis for revealing budget priorities" (p. 64).

I shall argue that the substance of the assertions that research has little bearing on the operation of schools and the patterns of support for research that follows from them are in egregious error; both must be altered if advances are to be made not only in educational inquiry but in educational practice.²

My argument proceeds in four parts. First, basic or theory-oriented research can have powerful effects on the operation of schools; I would challenge those who favor support for applied activity at the sacrifice of such studies to match the effects of their work against the effects that have accrued from the studies I shall cite. Second, far from contributing "only tiny 'bits'" of fragmentary and irrelevant knowledge, basic or theory-oriented research contributes to the broad conceptions and fundamental paradigms of human behavior; thus it provides the crucial contexts and guides for the substance and method of education. Third, these conceptions and paradigms enter into the preparation of school personnel and, in fact, as the paradigms are altered, preparation is altered (despite assertions to the contrary, ideas do have consequences). Fourth, I shall show the crucial practical, and ultimately applied, function the university performs through its basic research in formulating and pursuing apparently impractical problems: the definition of problems that might otherwise remain unpursued for being unformulated. It is this role that is being enfeebled by the unhappy confluence of financial stringency and the anti-intellectualism of funding agencies, which reflects their mistaken beliefs regarding the separation of basic research from practical application. And finally, I shall comment briefly on two competing views of the relation of theory, research, and practice and indicate the consequences of

² Although most of my examples will be drawn from psychological and developmental studies, this is not to suggest that these are the only types of research that sustain the argument: it is only that I am more familiar with these.

each view in education.

Before I turn to the argument proper, a word must be said about the term *basic research*. Basic research comprises studies in which the investigator formulates his own problem regarding a phenomenon or issue, and his aim is primarily to conceptualize and understand the chosen phenomenon or issue and only secondarily, if at all, to do anything about it. The work is theory oriented rather than action oriented. Although the distinction here is not foolproof, in this view basic research (like fine art) deals with "discovered problems" and applied research (like commercial art) deals with "presented problems" (Getzels & Csikszentmihalyi, 1967). All attempts at categorizing and naming the various kinds of inquiry are liable to ambiguity; there are always borderline cases between such categories as basic and applied, theoretical and practical, conclusion oriented and decision oriented. Perhaps the terms *theoretical inquiry* or *conclusion-oriented investigation* might be preferable to *basic research*. But the use of the latter is more prevalent for the studies I shall cite. In any case, what needs to be stressed is that no value judgment is intended regarding the necessity of one type of inquiry over another. Each has a distinctive function, each obtains stimulation from the other, and the flow of effect between the two is not unidirectional but multidirectional, not linear but organic. A pertinent fuller discussion of these issues is contained in a preceding publication by the National Academy of Education (Cronbach & Suppes, 1969).

1. EFFECTS ON SPECIFIC CLASSROOM PRACTICES

Despite the belief that basic or theory-oriented research has little effect on practice—a belief on which the sacrifice of basic research for other activities is founded—the fact is that basic research can have powerful effects on practice. Rather than argue the issue in abstract terms as has been done so often, I shall cite specific illustrative basic studies that have had manifest effects on our attitudes toward pupils, on leadership style, on classroom management, and on the nature of curriculum.

1.1 *Effect on Attitudes Toward Giftedness and Programs for the Gifted*

Consider, as a first instance, teachers' attitudes toward giftedness and gifted pupils (Getzels, 1968; Getzels & Dillon, 1973). Genius and insanity, it was held, are related and "great wits are sure to madness near allied," according to the familiar verse. There was hardly a belief regarding intellectual precocity and superior attainment that was more firmly fixed than that it was either a pathological condition in itself or was promoted by a pathological condition. The ancients believed that the genius was possessed by a god, muse, or demon. Early "scientific" support for the notion of intellectual giftedness as inseparable from personal deficiency

came from such work as *The Insanity of Genius and The General Inequality of Human Faculty Physiologically Considered* (Nisbet, 1891) and Lombroso's influential *The Man of Genius* (1891). This view was clearly expressed by Lange-Eichbaum (1932): "It may be a bitter pill for mankind to swallow, this recognition that in most instances genius and 'insanity' are inseparable" (p. 178). Jung (1954) spoke of the "narrow margin between a gift and its pathological variant" and the difficulty of determining in some cases "whether it is the gift or the psychopathic constitution that predominates" (p. 141).

It was within this context that intellectually gifted children, the so-called quiz kids, were thought of as "queer kids"—physically underdeveloped, emotionally maladjusted, socially immature, and morally suspect. Moreover, it was believed that unless their intellectual precocity was in some way restrained, these children would come to a bad end since "early ripe, early rot."

Terman's basic research (Terman, 1925; Burks, Jensen, & Terman, 1930; Terman & Oden, 1947, 1959) played a dominant role in altering these beliefs about gifted pupils, beliefs previously held without question by some of the most practical schoolmen. As is now well known, despite the then popular beliefs, Terman found that instead of being inferior, the gifted children were superior to other children not only intellectually but also physically, socially, emotionally, and morally. Moreover, their superior achievement in school, where they were typically two and sometimes four grades beyond their age-group, applied so generally to the different school subjects that it refuted another traditional belief that gifted children are usually "one-sided."

As for "early ripe, early rot," Terman's periodic follow-up of the original gifted children showed that they maintained and even increased their early superiority. The incidence of ill health, mortality, insanity, delinquency, and alcoholism was well below that of the general population of corresponding age. Close to 90% of the gifted group entered college, and 70% graduated; 30% of the graduates were awarded honors and about two-thirds went on to graduate work. The achievement of the 800 men at middle life (only a minority of the women had gone into professional careers) was equally impressive. Included among the men were 78 who had taken a PhD degree, 48 a medical degree, 85 a law degree, and 74 who were teaching or who had taught at the college or university level. By 1950, when the men were 40 years of age, they had published a total of 93 books (30 of them novels); 2,000 scientific, technical, and professional articles; more than 375 short stories, novelettes, and plays; and some 275 articles of a miscellaneous sort. All these numbers are many times as large as would be found for 800 men of corresponding age picked at random in the general population. And so another of the entrenched myths about gifted children (i.e., early ripe, early rot) was

found not to hold (Terman, 1954).

To be sure, others had held similar beliefs about these children. But Terman's basic research demonstrated what had previously been argued only as opinion. The growth of educational programs for the gifted student is directly traceable to this work; I was unable to find any text, curriculum, or program on the education of the gifted that does not make reference to the *Genetic Studies of Genius*. It is Terman, and not the producer of a prescription demonstration, or disseminator of an instant application, who is appropriately called the "father of the gifted child movement" in education. As Carroll (1963) pointed out some years ago, the attitudes we hold today toward giftedness and toward gifted pupils stem from Terman's basic research.

1.2 *Effect on Leadership Style and Classroom Environment*

Consider next the impact of the apparently simple little study by Lewin, Lippitt, and White (1939) on how children behave when faced with different kinds of leadership. The observations showed that behavior appeared to be a function of whether the leader was "autocratic," "democratic," or "laissez-faire." These are of course part of today's lore, and it is doubtful that those who use the terms give their source a second thought. The words *autocratic*, *democratic*, and *laissez-faire* have become part of the educational vocabulary as if they had been given by the good Lord at the Creation, instead of by a piece of basic empirical research in Iowa during the 1930s.

Within less than a decade after the study was reported, scores of treatises, textbooks, and school programs applied the findings to education. I counted no fewer than a dozen texts in educational administration alone deriving from this study, and even the physical conformations of classrooms were altered to provide for "democratic" rather than "autocratic" relations between teachers and pupils. We shall return to these subjects in greater detail in due course. Suffice it here to ask: Was there any educator—teacher or administrator—during the 1940s and 1950s who was not made aware of this study and of the effects of these "styles of leadership" on the "group dynamics" of the classroom and the school? Indeed, is there an educator who is not in some wise aware of them to this day?

None of this is to say that the vocabulary and point of view that were introduced by the study were necessarily as salutary as their adherents claimed. Nonetheless, although I do not wish to get into the silly argument about whether basic or applied inquiry is more influential, my contention being that even when they can be distinguished there is a symbiotic relation between the two and neither should be sacrificed for the other, I cannot help wondering whether dollar for dollar (if by no other criterion) many practical packages or prescriptions have had a

greater impact on the schools than this little basic inquiry.

I cannot help wondering, too, whether of the packages or prescriptions that had an impact there were not many explicitly or implicitly indebted to the concepts and findings of theory-oriented or basic research. Would, for example, the various programmed instruction and behavior modification packages, which the same funding agencies that deny the bearing of research on practice support with avidity, have been possible if Skinner's basic research on learning and reinforcement (with pigeons yet) had not been available? As one of the producers (Sherman, 1974) of the widely used programmed Personalized System of Instruction (PSI) explains, "Only as a sophisticated theory of learning developed, were we in a position to both avoid the crippling consequences of punitive systems and still maintain a standard of excellence. This required constructing adequate systems built on positive consequences. It is that theory of learning, the experimental analysis of behavior, that is the theory behind PSI" (p. 225).

1.3 *Effects on Curriculum*

Consider now, as another instance, the effect upon curriculum of Thorndike and Woodworth's (1901a, 1901b, 1901c) research on the doctrine of formal discipline or, more narrowly, on transfer of training—research that by no stretch of the imagination falls into the categories of development, demonstration, and dissemination. Nor did it produce any packages for instant application.

It will be recalled that according to the doctrine of formal discipline, the mind was viewed as though it were a body, and the training of the mind was seen as analogous to the training of the body. Just as muscular powers could be strengthened through arduous physical exercises like running, jumping, or lifting dumbbells, so the faculties of the mind could be strengthened through arduous educational exercises, exercises which need have no knowledge value in themselves.

Here, for example, is an excerpt from the prestigious *Pedagogical Seminary* at the turn of the century:

Intellectual training stands on very much the same basis as physical training. A man's physical nature can be trained by doing useful work or the exercises of the gymnasium, which have no value whatever except their effect upon the physical system of the performer. So one's brain system can be trained in studies that have a knowledge value for the individual, but also in those that have none. In both cases, of course, it is best that the gymnastics should be secured in the performance of useful work, as in this case two ends are gained at the same time; but as, perhaps, all kinds of work only partially develop one's physical powers, so that it is necessary to have recourse to gymnastics to complete the physical training, in the same way it may be necessary to have special exercise to develop particular brain functions, though such exercises have no knowledge value in themselves. (Hugh, 1898, p. 604)

This doctrine, which rested securely on tradition, authority, and common sense, was beyond questioning. It was used to justify Latin and Greek in the curriculum, since these provided manifestly arduous exercises. As a representative educational text of the period explained without hesitation:

The study of the Latin language itself does eminently discipline the faculties and secure to a greater degree than that of the other subjects we have discussed, the formation and growth of those mental qualities which are the best preparation for the business of life—whether that business is to consist in making fresh mental acquisitions or in directing the powers thus strengthened and matured, to professional or other pursuits. (Cited in Thorndike, 1913, pp. 360-361)

In the now-classic set of experiments, Thorndike and Woodworth tested the influence of exercises in technique, devoid of content, on one's ability to handle other problems—for example, the influence of training in estimating magnitudes on the ability to estimate magnitudes of the same general sort; the influence of training in estimating weights on the ability to estimate the weight of other objects; and the influence of practice in marking words with certain letters on the ability to mark words with other letters, misspelled words, and so on.² They found some carry-over, but less than expected. The general conclusion was that while there was some improvement, it was not due to any "mysterious transfer of practice" or "to an unanalyzable property of mental functions," but rather to a functioning of identical elements in the practice series and in the final test series.

The doctrine of formal discipline had of course been called into question before, notably by the Herbartians. But Thorndike and Woodworth's experiments had an enormous impact on education by demonstrating empirically that the mind just did not work that way. As one education text said within a decade of the experiments, "The two pieces of experimental evidence which have probably had the most influence in discrediting the doctrine of formal discipline and in emphasizing the opposite doctrine are those of James of two decades ago, and the more recent series inspired by Thorndike of Columbia University" (Colvin, 1913, pp. 215-216).

The Thorndike and Woodworth experiments were widely cited by the curriculum makers and contributed to profound alterations in the course of studies of the schools. To cite one instance for many, here is a curriculum article in the May 1909 issue of *Education* (Horne, 1909). The piece is entitled—and mark the title well, for it makes precisely the point of my argument—"The Practical Influence of the New Views of Formal Discipline." It begins by giving a bibliography of the work on the "new views," including the work by Thorndike, and summarizes the experiment's conclusion as:

Present study helps us in the later situations in so far as common features are involved. . . . [That is] as there is no exact correspondence between mathemati-

cal and life situations, a good mathematician may be a poor reasoner in nonmathematical situations in life, and a good reasoner in such situations may be a poor mathematician. These statements are not opinions, but experimentally demonstrated facts. (p. 616)

From these "experimentally demonstrated facts" rather than the former "opinions" comes the rationale for the new curriculum:

Now, as the educational aim is the unfolded and capable mind in the concrete social and natural situations of life, and as we are efficient in those situations in proportion as we have developed ourselves earlier in similar situations, it follows that those subjects have the greater educational value which have the greater number of identical elements with the situations of life. There are no subjects which give "all 'round mental training," for no such training is possible. There is no small, select number of drill subjects which will so discipline the mind that it can meet the problems of life with greater efficiency. (p. 616)

The writer adds, "The Latin grind in the secondary school is left to justify itself as a matter of formal discipline. But now this ground is swept from under it. We have nothing to anticipate in consequence but a necessary decline in the percentage of students who study Latin" (p. 618). Compare this with the earlier remarks regarding the study of Latin as the "best preparation for the business of life" (no bearing of research, of basic research, on the actual operation of schools?).

Thorndike and Woodworth's basic research contributed not only to altering the role of Latin and Greek in the curriculum but to the very conception of the mind as well. Good reasons for learning Latin and Greek remained, but their use as dumbbells to strengthen the muscles of the mind was not among them. I would urge anyone who doubts the contribution of basic research to the practice of education to compare several curriculum tests or guides on the place of Latin and Greek as school subjects before and after the Thorndike-Woodworth studies.

Consider as a second and more contemporary example the impact on the curriculum of the recent basic research on cognitive development. Indeed, consider only the impact in this area of the research on the conservation of volume.

Who would have thought that the investigation of children's ability to tell whether the volume of water in one glass tube is the same as that in another glass tube would one day be a major influence on the curriculum and methods of our schools? To use only science teaching as a case in point, an entire issue of the *Journal of Research in Science Teaching* in 1964 was devoted to the influence of Piaget's studies of cognition on the creation of new curricula and methods in science teaching. More specifically, Karplus (1964) describes the intent and construction of his Science Curriculum Study material as that of fostering the development of children's thinking to the formal operational level. From contacts with teachers over a period of years, he concluded that the "stimulus-response" method of teaching science had a "powerful hold" on them, and that two

practical strategies were needed to counteract its harmful effects. The source for both strategies was in basic research on cognitive development.

One effort was to lead teachers toward a greater understanding of the different stages of intellectual development so that science material and instruction would be calibrated to stages. This was done by introducing the teachers to Piaget's experiments. Karplus (1964) writes, "I think the conservation of volume study is one of the most impressive demonstrations that any teacher can carry out with children of age five, six, or seven. Those teachers to whom we have suggested this study and who have tried it have been amazed. They would not have believed the results ahead of time" (pp. 236-237).

The second effort was to produce an actual curriculum and set of materials that would help children reach the formal operational level. Karplus describes one of the topics in the curriculum as "Material Objects." Here children at the kindergarten and first-grade levels become aware of material in the environment and the conservation properties the material objects have. Another topic for the first grade is "Variation and Measurement." These tasks are more difficult because measurement involves a conservation of an object's property with respect to transformations which may involve displacements from one point in space to another. Citing Piaget's development of these concepts, Karplus writes, "The ruler's invariant length has to be recognized; otherwise comparing one ruler to many objects when measuring them does not particularly make sense" (p. 238).

Over and above the specific curriculum topics and materials which derive from the basic research on conservation and invariance, Karplus writes regarding the more general problem of educational method:

I believe Piaget has pointed out that in the natural development of a child his own initiative to shape his explorations and to decide what he will do is dependent upon the child—the child decides. I think that this economy of decision is a significant element which determines the quality of the learning that occurs. The question we must therefore face is, how much help can one give the child without depriving him of his decision-making initiative to such a large extent that he would never recognize when a particular idea really works out? The Science Curriculum Improvement Study's compromise is to have the teacher provide the construct, but to leave ample opportunity for the children to apply the construct to many cases in their own manipulation. (p. 239)

1.4 *Other Basic Research*

Are the instances I have cited merely exceptional cases, and am I stacking the studies to make my point? I believe not. I could have cited many more, but I do not know how many instances will be persuasive to those who cannot already think of their own favorite ones.

I could have cited Guilford's basic research on the structure of intellect,

which led to the notions of divergent thinking as the substratum for originality, increasingly a part not only of the assessment of children's ability but of the curriculum objectives in many schools; or Hebb's basic research on the effects of sensory deprivation, which raised important questions about the role of the early life of disadvantaged children on their later performance in school; or Clark's basic research on the self-concepts of black and white children, used by the Supreme Court in its desegregation decision, and underlying much of the thinking about integration; or to take a final, rather different type of example, one out of keeping with my self-imposed restriction to psychological research, I could have mentioned Theodore Schultz's basic research on the economics of education, which raised the crucial question of whether schools consume or produce capital.

To be sure, the indicated studies did not just happen. The flow of influence was not only from the theories of the academy and the experiments of the laboratory to the practices of the classroom and the procedures of the school; it was from the practices of the classroom and the procedures of the school to the theories of the academy and the experiments of the laboratory as well. And of course there are other influences of a social and political order involved in whether a change indicated by research will or will not be adopted. Even if adopted, the change may not be an improvement. Moreover, to cite instances of research that did have an impact is not to gainsay there is research—perhaps the greater proportion of work—that is trivial and without influence on practice. But the point remains: Despite the incessant cries to the contrary by schoolmen, funding agencies, and often educational researchers themselves, theoretical inquiry and basic research can and do have practical consequences on the operation of schools.

2. EFFECTS ON BROAD EDUCATIONAL POLICY

I come to the second issue, the charge that even when it is productive, basic research—or “merely theoretical” research, as it is sometimes scornfully referred to—provides only “tiny ‘bits’ of knowledge” and, as such, is irrelevant and without influence on the broad problems of education. Nothing, it seems to me, can be further from the case.

The significant impact of research comes not piecemeal, study by study, technique by technique, practice by practice. It comes cumulatively, and, in this sense, the preceding section citing separate studies in conceptual vacuums underrepresents the effect of research. It is precisely basic or theory-oriented study in its cumulative impact that alters the general conceptions, and ultimately what Kuhn (1962) calls the paradigms, of the human being and his condition, which in turn have an influence on the broadest educational problems.

2.1 *Conceptions of the Learner and Architecture of the Classroom*

Consider, as one example in this regard, the alterations in architecture of the classroom, symbolic of more profound but less visible educational policies, of our conceptions of the "learner"—conceptions influenced by, if not founded in, theory-oriented research.

Almost within sight of my office are four school buildings. In one, dating from the turn of the century, the spaces called classrooms are rectangular, the pupils' chairs are bolted to the floor, and the teacher's desk is front and center. In the second building, dating from the 1930s, the classrooms are square, the pupils' chairs are movable, and the teacher's desk is out of the way in a corner. In the third building, dating from the 1950s, the classrooms are also square but the pupils' desks are trapezoidal in shape so that when placed next to each other they form a circle, and the teacher's desk has vanished. In the fourth building there is a classroom constructed in the 1970s that is four times the size of the ordinary classroom and has no teacher's or pupils' desks but is filled instead with all manner of odds and ends from finger paints to Cuisenaire rods. If one were not told it was a classroom, this space might be mistaken for an overgrown playroom or a warehouse full of children's paraphernalia. I shall refer to the spaces in the first building as the "rectangular" classroom, in the second as the "square" classroom, in the third as the "circular" classroom, and in the fourth as the "open" classroom.

Since architectural space or form presumably follows from the function for which it is intended, and since the function—teaching and learning—is presumably the same for all four classrooms, why are there then these obvious differences among them? More specifically, is there some reason for the succession of the different forms, or was the succession merely accidental? Indeed, there is the more fundamental question: Why does any classroom have a particular architectural form?

To be sure, if one looks into the classrooms in operation, the differences often seem to disappear. The movable chairs are placed in straight rows and the trapezoidal desks are lined up behind one another rather than joined in the circle for which they were designed. But that is not the problem I am addressing. Despite usage, the design of the altered classroom has obviously been *not* to have straight rows. The problem is how to account for the standard straight-row classroom represented by the original bolted school chair and for the design (if not always actuality) of the altered classrooms represented by the changing school chair.

The form of the standard straight-row classroom is usually attributed to the 19th-century tradition of building and architecture, the physical requirements for lighting and ventilation, and the pedagogy of the time which emphasized pupil discipline. The changes in the subsequent forms of the classroom are attributed to alterations in these architectural, physical, and pedagogical considerations (see Sommer, 1969; Educational

Facilities Laboratories, 1960).

I do not wish to minimize these influences, nor the effect of social and political forces. But there is another influence, albeit a less manifest one, that may ultimately be as decisive a determinant of the classrooms we build, and this is the conception of the learner. Just as it is impossible to understand the differences between the forms of the Catholic cathedral and the Congregational prayer house, or the Lutheran church and the Quaker meeting room, as spaces for worship by looking only at the traditions of building and architecture, the physical requirements for habitation, or even the materials and methods of the litanies without considering also the different visions they reflect of man as worshipper; so it is impossible to understand the differences in the form of the classroom as space for learning without considering the different conceptions of the child as learner. And, to repeat, the differences were influenced by, if not founded in, basic or theory-oriented research.

2.2 *Empty Organism: Rectangular Classroom*

At the turn of the century, a dominant, although by no means only, conception of the child as learner was that he was ideationally an "empty" organism responding to stimulation and learning primarily by trial-and-error when specific responses were connected to specific stimuli through the mediation of pleasure or pain. The organism itself, it was believed, would not learn or think if it was not impelled into these activities by internal drives like hunger and thirst, or by externally applied motives like reward and punishment.

It was the so-called Associationist or Connectionist period in the conception of the learner and of learning. In the words of the major proponent of this view:

Both theory and practice need emphatic and frequent reminders that man's learning is fundamentally the action of the laws of readiness, exercise, and effect. He is first of all an associative mechanism working to avoid what disturbs the life-processes of certain neurones. If we begin by fabricating imaginary powers and faculties, or if we avoid thought by loose and empty terms, or if we stay lost in wonder at the extraordinary versatility and inventiveness of the higher forms of learning, we shall never understand man's progress or control his education. (Thorndike, 1913, p. 23)

Researchers in the learning laboratory were "connecting" desired stimulus-response bonds of animals in cages, and teachers were "connecting" desired stimulus-response bonds of children in classrooms.

This description is of course an oversimplification and to some extent a caricature, as it must necessarily be in so brief a sketch. And I need hardly add that in pointing to this aspect of the turn-of-the-century view of the learner, I do not intend to derogate the historic achievement of the Associationist conception of learning. Nonetheless, the essential point holds: The prevailing psychological conception of the learner was that he

was more or less an ideationally empty organism associating stimuli and responses through the operation of rewards and punishments under the control of the teacher. Both the learning stimulus (what was to be learned) and the learned response (what was actually learned) were believed to be determined by the teacher. In effect learning was "teacher-oriented."

It was no accident, then, that the prevailing methods and materials and even the school buildings and classroom arrangements were teacher-centered. Typically, the classroom was rectangular in form with the teacher up front—sometimes even on a dais or platform—and the pupils in chairs rigidly fastened to the floor in straight rows facing forward so that they could not turn away from the presumed source of their learning experience: the teacher.

Given the particular conception of the learner, what could be a more eminently sensible and practical learning environment? Indeed, there is a letter by John Dewey dating from this period in which he complains that when he was trying to furnish his new Laboratory School according to his different conception of the learner and of learning, he was unable to find any other kind of classroom furniture (McCaul, 1973).

How then did we evolve from this apparently sensible and practical rectangular classroom and all that this symbolizes in curriculum and instructional method to the more recent, apparently equally if not more sensible and practical, but almost opposite, classroom—with no teacher's desk, to say nothing of a dais, and the pupils' movable chairs all over the place (and all that *this* symbolizes in curriculum and instructional method)?

This transformation did not occur because someone suddenly happened to have a bright idea that a change in the physical appearance of the classroom might be a good thing, or because it was readily demonstrable that children in movable chairs were easier to discipline than children in fixed chairs (if anything, the reverse is more likely), or, more to the point here, because there was any specific practical investigation showing that children in the new movable chairs learned better than in the old, fixed chairs. As I have already pointed out, there were unquestionably many factors contributing to this transformation, including the general progressive movement in social and political thought (Cremin, 1961); and the flow of influence was surely not only from the laboratory to the classroom; it was from the classroom to the laboratory as well. But an important factor was the modification in our conception of the human being, and more especially of the human being as learner. And the important research, the work that was of fundamental influence, was not so much the research directed at specific practices but the theory-oriented inquiry directed at understanding the human being and his behavior, that included the processes of learning and teaching.

2.3 *Active Organism: Square Classroom*

It became evident that the earlier conception did not account for all or even most of the behavior observed in the learning situation (Segal & Lockman, 1972). The individual did not seem to experience what he was to learn as discrete stimuli. Instead he saw apparently discrete stimuli as "belonging together," as making a configuration or gestalt. In the learning laboratory, three discrete dots placed in certain relation to each other were seen not as three separate dots but as a single triangle; two open brackets facing each other were perceived as one rectangle; even chimpanzees learned by perceiving gestalten. And in the classroom it was recognized that two pieces of information given separately by the teacher were often brought together by the learner himself into a single idea not given by the teacher in the separate pieces. Experiments in the learning laboratory were dealing with problems of "perception" and "pragnanz"; and teachers in the classroom, with problems of "insight" and "closure."

Moreover, personality theory was postulating that the human being was not psychologically an empty organism. On the contrary, he was a tumultuous bundle of needs, interests, values, persuasions, projections, repressions, and conscious, preconscious, and unconscious psychic forces that determined his behavior. What the learner would and would not respond to, what he would and would not learn, were seen as determined by these forces. In research studies, when individuals were asked to recognize equally ambiguous words, some of which were congruent and others incongruent with their values, they tended to recognize the former more rapidly and accurately than the latter (Postman, Bruner, & McGinnis, 1948); when individuals were required to learn material favorable and unfavorable to their persuasions, they later recalled what was favorable and forgot what was unfavorable (Edwards, 1941; Levine & Murphy, 1943).

The conception of the learner as an empty organism was transformed into the conception of the learner as an active organism. Learning was conceived as not only a connective process but as a dynamic cognitive and affective process as well. From this point of view, the learner—not the teacher—was the center of the learning process. Experiments in the laboratory were concerned with the relation of the learner's personality to his learning; and teachers in the classroom, with the relation of the learner's needs to his adjustment in the school situation. Learning became more "learner-centered."

It was no accident, then, that the character of at least the ideal classroom took on a new conformation. The teacher-oriented classroom became the pupil-centered classroom. The teacher's desk was shifted from the front of the room, out of the way, to the side; and the pupils' rigid chairs in the straight rows in a rectangular-shaped space, which had

seemed so sensible and practical under the older conception of the learner, became primitive if not forbidden objects. They were replaced by movable chairs that could be shifted at will in the new square-shaped classroom according not only to the requirements of the teacher but, even more importantly, to the needs of the pupils. And this, it must be repeated, not because of any single piece of applied research that provided a practical package for doing something, but through the accumulation of basic or theory-oriented research that contributed to the understanding of the human being and modified the prevailing view of the child as learner. The altered construction of the classroom was at least, in part, a function of the altered conception of the learner.

2.4 *Social Organism: Circular Classroom*

If the conception of the learner in the initial period was predominantly as an empty organism, and in the next as an active organism, in the period that followed it was as a social organism. The beliefs regarding the nature of the learner in the first period drew largely from the Associationist view of the human being, in the second from the personalistic and Gestalt views, now they drew also from the emerging social-psychological and field-theory views.

The child as learner was envisioned as a social organism, and learning was perceived as occurring in a field of interpersonal actions and reactions, each person in the classroom serving as a stimulus for every other person. As has already been pointed out, it is hard to overestimate the impact on the classroom of the leadership and group climate studies by Lewin and his associates beginning in the late 1930s, which were given further impetus by the ideological issues of World War II (Getzels, 1969). Innumerable treatises, textbooks, and programs applied the ideas to the classroom, and such terms as "group dynamics" and "autocratic," "authoritarian," "democratic teaching" became part and parcel of the language of education. As Wallen and Travers (1963) say, "We suspect that the Lewin, Lippitt, and White (1939) study is better known to education students than any other single piece of research" (pp. 474-475). Experiments in the laboratory were concerned with such matters as "interpersonal cohesion" and "communication networks"; and teachers in the classroom were concerned with "sociometric structure" and "group processes." Learning became more "group-centered."

Changes concomitant with the conception of the learner as a social organism were introduced into the curriculum and into the physical conformation of the desirable classroom, and of a good many real ones. If the child is primarily a social creature, then the objectives and strategies of his education should be social as well as cognitive; his sociometric status is as important as his intellectual status. And if learning itself is a social or group process, then a circular or group-centered classroom where

everyone faces everyone else (as once they had been forced to face only the teacher), or at least furniture like conference tables or a circular seating arrangement (which will serve the same purpose), is the most sensible and practical learning environment (as the rectangular straight-row arrangement had been before). These in fact became the favored curricula and learning environments, as the architecture of the classrooms built at the time will attest. And again, these alterations did not occur because of any particular study that contributed a bundle of prescriptions for immediate application but because of theory-oriented studies that contributed to a conceptual and strategic transformation of a more general order.

2.5 *A Changing Paradigm*

What about the present? How did we get from the straight-row classroom or even the circular classroom to the so-called open classroom? I have been emphasizing the dissimilarities among the preceding conceptions of the learner. There was, however, a critical similarity among them. The preceding views of the learner had this underlying paradigm in common: They were founded in a combination of the homeostatic model of self-maintenance and the drive-or-tension-reduction theory of behavior (Getzels, 1964; Hunt, 1960; White, 1959).

According to this paradigm, the organism's natural state is conceived to be equilibrium, and the organism always acts in such a way as to return to this optimum natural state. Hunger produces tensions driving the organism to seek food, and it ceases to seek food when it has eaten; fear produces tensions driving the organism to seek cover, and it ceases to seek cover when the danger is over; lack of closure produces tensions driving the organism to seek closure, and it ceases to seek closure when the gap is filled; conscious or unconscious conflict produces tensions driving the organism to seek resolution, and it ceases to seek resolution when the conflict is resolved. With respect specifically to learning (or thinking, problem solving, and other forms of intellectual activity), the organism is similarly said to be driven (it is called "motivated" in education) as it is in any other form of behavior. As Hilgard (1948) describes this position, "Without drives the organism does not behave and hence does not learn" (p. 78). Learning always involves the reduction of a drive, a decrease in tension and stimulation.

There has been a growing discontent with this paradigm of behavior, at least as it is applied to learning and other forms of intellectual activity. To be sure, learning, thinking, problem solving, and intellectual exploration may be means to reducing certain drives. But they are not *only* that. They may also be ends in themselves, the organism acting to *increase* as well as to decrease stimulation.

One of the most obvious features of animal behavior, as Robert White

(1959) points out, is the tendency to explore the environment. In his words, "Cats are reputedly killed by curiosity, dogs characteristically make a thorough search of their surroundings, and monkeys and chimpanzees have always impressed observers as ceaseless investigators" (p. 298). In a series of illuminating experiments, Harlow (1954) and his co-workers showed that monkeys who are viscerogenically sated, that is, not hungry, thirsty, or otherwise driven by any recognizable primary drive, will learn to unassemble a puzzle of some complexity for no other apparent motive than, as Harlow says, "the privilege of unassembling it."

The behavior of the animals in learning to solve puzzles reduced no recognizable drive, decreased no identifiable tension or stimulation. Far from trying to reduce tension or stimulation, the animals seemed to seek stimulation; once they had learned to solve one puzzle, they seemed to become "bored" and turned spontaneously to solving a different puzzle.

Discontent with the drive-reduction paradigm has not been restricted to animal behavior. Piaget (1952, 1954), Hunt (1960), Schachtel (1959), and White (1959), among others, point out that the central fact in the growth and development of children is not hunger satisfaction or thirst satisfaction or some other so-called primary drive satisfaction; it is the opportunity for effective interaction with the environment as manifested in the child's curiosity and exploratory activity. A child all of whose bodily needs have been satisfied does not remain at rest; it is just then that the child's play is most active and constructive.

Moreover, a conspicuous portion of adult life is devoted to activities that can be classified only as "fine art," "pure science," "adventure for its own sake," activities that cannot be said to contribute to bodily maintenance through satisfying primary drives and attaining states of rest (Berlyne, 1966). Donald Hebb (Hebb, 1949; Hebb & Thompson, 1954) points, in addition, to common activities like working at riddles, reading mystery stories, or traveling to strange places, all of which provide pleasure not by reducing some bodily drive but by raising the level of excitement. The normal person, he argued, needs sensory stimulation for his well-being just as he needs food or shelter, and he will seek out such stimulation if it is not readily available.

Bexton, Heron, and Scott (1954), as one example for others, put this observation to experimental test. They paid subjects "to do nothing." The subjects were well fed and comfortably housed. But there was one condition. Stimulation was minimized; they could not see, hear, touch, or communicate with anyone. At first the subjects merely relaxed or slept. Soon, however, they became restless, could not sleep, and exhibited random behavior as if looking for something to do, problems to work on. The situation became so unbearable that despite the satisfaction of all the bodily needs and the pay for presumably doing nothing, it was impossible to keep the subjects in the experiment for more than two or three days.

On leaving, they reported feelings of fatigue, nausea, and mental disorganization. The human being seems to require an optimal level of stimulation and an opportunity to seek out problems for solution if he is to function effectively.

In short, recent research directed toward the *understanding* of behavior, that is, primarily toward theory rather than practice, is suggesting a change in the conception of the human being and with this a reconsideration of the character of his education as manifested by the newer classroom forms and teaching strategies.

Put most briefly, this seems to be the core of the changing paradigm of the human being: In addition to the familiar viscerogenic drives like hunger and thirst (on which so many assumptions about the learner are based, when they are not based on the hungry animal in a maze) that are satisfied by satiation, there seem to be as yet ill-defined neurogenic and psychic needs that are gratified by stimulation—needs for excitement, novelty, sensory variation, and perhaps above all the challenge of the problematic (Getzels, 1964).

2.6 *Stimulus-seeking Organism: Open Classroom*

From this point of view, the human being as learner is not like a mechanical calculating device, or an information processing device, that lies idle until it is sparked by some drive, stimulus, conflict, problem, or teacher goad; rather, the human being as learner is a vital entity perpetually in motion (even when apparently at rest) that must be given an effective opportunity to remain in motion if it is to function and remain vital. Human beings need not be driven to explore, to think, to learn, to dream, to seek out problems for solution; they are intrinsically constituted to do just this. The learner is not only a problem-solving and stimulus-reducing organism but also a *problem-finding* and *stimulus-seeking* organism.

It is admittedly dangerous to encapsulate complex phenomena in catch-phrases, but if the ideal classroom that emerged from the conception of the learner as an empty organism was teacher-centered, the classroom that emerged from the conception of the learner as an active organism was pupil-centered, and the classroom that emerged from the conception of the learner as a social organism was group-centered, the altered classroom that seems to be emerging from the conception of the learner as a stimulus-seeking organism may be characterized as "inquiry-centered."

The so-called open classroom, as one instance, is isomorphic with the stimulus-seeking paradigm of the learner and the inquiry-centered view of a desirable learning environment. Barth (1972), to cite a specific case in point, describes his attempt to institute such a classroom and gives as the first assumption underlying this type of learning environment: "Children are innately curious and will explore without adult intervention" (p. 18).

He gives as the second assumption: "Exploratory behavior is self-perpetuating" (p. 19).

Just as the appropriate spatial organizations of the preceding classrooms were compatible with the preceding conceptions of the learner, so an altered spatial organization of the classroom follows from this altered conception of the learner. This does not mean that all that needs to be done is to fill a space with odds and ends that *might* be interesting to children and let nature take its course. On the contrary, the arrangements often require more, rather than less, teacher skill and attention. In Barth's words,

The teacher in the open classroom organizes his classroom not to produce optimal conditions for transmission of knowledge but to extend the range of possibilities children can explore. Children's desks are often removed from the room, and tables and chairs are left. . . . Space within the the classroom is divided, often by movable screens or furniture, into an "interest area." . . . While in the traditional classroom the child learns at this desk, in the open school *the locus of learning is where something of particular interest for the child happens to be.* (p. 74; italics in original)

If one accustomed to the older classrooms were not told this was a classroom too, the space might easily be mistaken for an overgrown playroom or a warehouse full of children's paraphernalia. Indeed, if this classroom is judged by the conception of the learner as a psychologically empty organism who learns only by associating stimuli and responses through the operation of rewards and punishments, then it cannot look like a classroom and seems an inappropriate place for learning. But if the arrangements are judged by the revised conception of the learner as a stimulus-seeking organism, what did not look like a classroom looks very much like an appropriate environment for children to learn, appropriate for permitting the search for stimulation they need to reason, think, and learn. The important word here is *appropriate*—appropriate as judged within the paradigm. It is now the more traditional classroom with the straight rows and rigid chairs, which had once seemed so appropriate within its paradigm of the learner, that seems the inappropriate environment for learning, whether it is so or not.

When one visits these various classrooms in operation, one is confronted with a striking anomaly. The movable chairs are often placed in straight rows, the trapezoidal desks are lined up behind one another, the open classroom is partitioned into separate closed-off spaces. In one renovated classroom I visited, the movable chairs were placed exactly on the bolt marks left by the fixed chairs they had supplanted, and woe to the child who moved his chair off the bolt mark. Despite the usage, the design of the altered classroom and furniture had obviously been *not* to have straight rows. Why then the reversal? In a paradoxical way, this is a rather nice demonstration of the impact of conception on practice in the class-

room. The usage conformed to one conception of the learner, or at least to the methods the teacher had learned within that conception; the design of the classroom conformed to another conception. The discrepancy between usage and design is a reflection of differences in the conceptions and cannot be understood apart from these.

This is as far as we need pursue our illustration of the organic relations of basic research, paradigms of the human being, and the operations of classrooms. I need hardly add that the periods to which I referred were by no means as sharply defined as the schematic account unavoidably implies, the differences between the periods residing in the relative emphasis rather than the all-or-none presence or absence of one view or another of the learner and of one operation or another of the classroom. The catch phrases "empty," "active," "social," and "stimulus-seeking" for the learner, and "rectangular," "square," "circular," and "open" for the classroom, are no more than shorthand designations to be used with caution. Furthermore, there is the obvious principle that to point to change is not to imply benefit or progress.

Despite these qualifications, conceptions of the learner—conceptions having had their source, at least in part, in basic research—do have an impact on what goes on in the classroom. Far from providing only bits of knowledge irrelevant to the operation of schools, basic or theory-oriented research affects our paradigms of human behavior and thus implicitly or explicitly influences the forms and practices of the classroom.

What Keynes (1964) said regarding the power of so-called academic theories or ideas in economic affairs applies to educational affairs too:

The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back. I am sure that the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas. Not, indeed, immediately, but after a certain interval; . . . soon or late, it is ideas, not vested interests, which are dangerous for good or evil. (pp. 383-384)

3. EFFECTS ON PROFESSIONAL PREPARATION IN EDUCATION

I turn now to the third point. Not only do basic research and paradigms have an impact on practice but also, whether the practitioners know it or not, on professional preparation for the schools. As the paradigms are altered, preparation is altered. To be sure, the influence on preparation is not as direct as the influence of, say, Skinner on programmed instruction, or as visible as the changes in conceptions of the

learner on the form of the classroom. The influence enters more subtly and tortuously, but no less certainly.

3.1 *The Preparation of Educational Administrators*

As the case instance, I shall use the preparation of educational administrators. The issue to be addressed may be stated in specific terms by reference to the textbooks in this field. The foreword to the *first* edition of perhaps the most widely used text in educational administration says, "We feel that serious examination of and thought about one's own experiences and those of other people dealing with similar problems is of vital importance to anyone who wishes to gain a real understanding of the administrative process" (Campbell, Corbally, & Ramseyer, 1958, p. vi). The foreword to the *fourth* edition of the same text a dozen years later altered this to: "The book is built on the conviction that one develops an understanding of educational administration as one examines *concepts* and relates them to actual experiences" (*italics added*; Campbell, Bridges, Corbally, Nystrand, & Ramseyer, 1971, p. xiv).

How may this change in the preparation of administrators—and as we shall see, it was a crucial change—from reliance only on the *experiences* of already practicing administrators to the study also of the *concepts* of administration be accounted for? Or, more generally, how is one to account for the alteration in the content and authorship of two successive and widely used *Yearbooks* on educational administration by the National Society for the Study of Education, one in 1946 (Henry, 1946), the other in 1964 (Griffiths, 1964)? Of the nine persons responsible for the first volume, five were practicing state school officers or school superintendents; of the 15 persons responsible for the second volume, not one was. They were instead mainly from the administrative, organizational, and behavioral sciences.

3.2 *Scientific Management*

During the first quarter of the century or so, the predominant emphasis of administration was on "scientific management." This view held that excellence in management was "knowing exactly what you wanted men to do, and then seeing to it they do it in the best and cheapest way" (Taylor, 1911, p. 21). Among the principles of administration were: Management should take responsibility for planning the work from the worker, it should measure production effort by time and motion study, and it should set incentives before the worker in order to stimulate desirable effort on his part. Numerous treatises and textbooks—those by Bobbitt, Cubberley, Strayer, and Reeder—applied the principles to educational administration. Cubberley, for example, in his enormously influential text *Public School Administration* (1916), stated that the administrator must plan the educational policy but "the details of this policy he may find it wise to keep to himself" (p. 145).

This view of administration was not the aberration of deluded or self-seeking men or the result of a perverse accident. The principles, however untoward they may seem to us today, were consistent with the values of the time and with the prevailing conception of the human being and human behavior that determined the relations between pupil and teacher, teacher and administrator.

As we have already seen, the dominant conception of the human being was that he was by nature an empty organism responding only to pleasure and pain, reward and punishment. The person himself would do nothing unless impelled by the carrot of incentive or the goad of authority. It was no aberration or accident that the classroom of the time took the form it did with the teacher in command up front, the pupils submissively in rigid rows facing him, and a gold star or the birch dangled before them, since it was presumed that the child would not think or learn or do anything useful without these measures.

Given this conception of human nature, what could be a more sensible and practical classroom structure and pupil-teacher arrangement? And given the same values and conceptions of the human being, what could be a more sensible and practical administrative structure and teacher-administrator arrangement than one similarly based on the primacy of administrator initiative and teacher submission and the role of incentive and authority as advocated by scientific management?

3.3 *Human Relations Management*

The scientific management view of administration was predominant during the first quarter of the century; what came to be called the "human relations" view developed during the second quarter. This view held that human relations were the "warp and woof" of an organization and that everyone in it from those in the most menial to those in the most exalted positions should "evoke each other's ideas, come to see each other's viewpoints, and integrate their viewpoints in pursuit of the common goal" (Metcalf & Urwick, 1940, p. 14).

Support for the conception of management and administration based on human relations came from the studies by Elton Mayo and his colleagues, notably F. J. Roethlisberger and William J. Dickson (1939). From 1923 to 1932 they performed the now famous series of experiments at the Hawthorne plant of Western Electric, culminating in the volume *Management and the Worker*. These experiments have been liable to criticism, but whatever the merit of the experiments themselves, one must recognize their impact. They were taken to show that the workman was not an empty organism; what went on inside the workman and between him and the manager was more significant for production than other factors, even in the most scientifically managed and rigorously job-analyzed work situation (Getzels, Lipham, & Campbell, 1968, 33ff.).

Educational administration was influenced by the Hawthorne studies

and even more by Lewin's basic inquiry (Lewin et al., 1939) into different styles of leadership and group dynamics to which we have already referred. The results regarding the effect of democratic and autocratic leadership were applied not only to pupil-teacher but also to teacher-administrator interactions. There was an outpouring of treatises and textbooks for the preparation of school administrators from this point of view, often with direct reference to the basic Lewinian conception and studies. Here, for example, is a sampling of titles: *Practical Applications of Democratic Leadership*, *Human Relations in Educational Organization*, *Democracy in School Administration*, *Improving Human Relations in School Administration*.

Typically, the last-named volume (Yauch, 1949) describes the "discovery that under domination the quality of work deteriorated greatly when the autocratic leader was not present, while the quality of work in democratic climates continued at approximately the same levels" and concludes, "Apparently, in terms of the environment in which individuals work, it is better to let groups make their own choices, whether leadership is present or not, than it is to control the lives of others" (p. 40). Implicit in all the texts and stated explicitly in this one was the admonition: The primary responsibility of the educational administrator is to facilitate interaction in the faculty group so that "all individuals affected by any decision should have a share in determining its character and form" (p. 16). This is a far cry from scientific management's principle of separation of planning from performance and Cubberley's admonition to the prospective school administrator that he would be wise to keep his plans and policies to himself.

Just as it would be shortsighted to look on the scientific management view as the aberration of deluded or self-seeking men, so it would be shortsighted to look on the human relations view as only the inspiration of more sophisticated and virtuous men. Like scientific management, the human relations view was also consistent with the prevailing social forces, values, and conceptions of human nature—but an altering system of values and conception of human nature, to say nothing of social forces.

In values, the turn of the century's work-success ethic was giving way to an ethic of sociability; the American ethos shifted from Main Street to Madison Avenue and the model for our young from the hardworking Horatio Alger hero to the affable young man in the grey flannel suit. Future-time orientation was giving way to present-time orientation; the former national slogan "a penny saved is a penny earned" was replaced by the new slogan "buy now, pay later." Competitive individualism was giving way to adaptive conformity; "inner direction" was transforming to "other direction," self-made man to organization man. And Puritan morality was giving way to moral relativism; morality became a statistical rather than an ethical concept and was taken to be whatever the group one belonged to took it to be (Getzels, 1972).

The dominant conception of the human being was being transformed from an empty organism to a dynamic social organism. In a sense, Watson's and Thorndike's image of human nature was replaced by Freud's and Lewin's. The authority-centered classroom with its rigid straight rows where everyone could face only the teacher became at least ideally the group-centered democratic classroom with the movable chairs in circular arrangements where everyone had to face everyone else as before they had been forced to face only the teacher. The teacher-preparation texts that urged new teachers to be "democratic" toward their pupils were paralleled by administrator-preparation texts that urged new administrators to be "democratic" toward their teachers. There was more than an accidental likeness between the teacher's management of pupils through "sociometry" and "group dynamics" and the administrator's management of teachers through "consensus" and "human relations."

Although the principles of scientific management and of human relations were often antithetical, their assumptions for the field of administration had much in common. Administration was conceived of as a vocation or trade, not as a field of study; the orientation was primarily technological and prescriptive, not theoretical and analytic. The texts espoused principles of what an administrator should do, not concepts he needed to understand.

3.4 *From Vocational Principles to Theoretical Concepts*

At about mid-century a different view arose, both of administration and of preparing administrators. If scientific management with its stress on impersonal authority was the thesis of the first period, and human relations with its stress on personal interaction the counter-thesis of the second period, this third period was concerned with a theoretical synthesis of the two. Administration, it was held, is not only a vocation but should properly also be a field of study; not only the technologies but the concepts of administration must be dealt with. The administrator must not only know what to do but must understand why he is doing it.

An early statement of this view was Herbert Simon's *Administrative Behavior* (1945). Simon argued unequivocally that the accepted principles of administration were "little more than ambiguous and mutually contradictory proverbs" (p. 240). A different approach was needed, one that would establish a consistent and useful administrative theory. Such an approach was to be found in shifting the emphasis from principles to *concepts* of administrative behavior. These concepts must deal with the relations between the rationality of the individual and the constraint of the organization as they affect administrative decisions. As Simon put it, "Administrative theory must be concerned with the limits of rationality, and the manner in which organization affects these limits for the person making a decision" (p. 241).

Bakke, Argyris, and the Yale research group in management, as another instance, also argued for the primacy of theory: "Every leader of organized activity whether he be a manager of a business, an official of a union, an executive of government, a pastor of a church, a commander of a military force, a president of a university, or a father or mother of a family, bases his decisions and actions on an organizational theory" (Bakke, 1952, p. 1).

They too took as the central issue in administration the synthesis of the personal or individual and the impersonal or organizational requirements in order to maximize the "actualization" of both. As they argue, when an individual and an organization come together they do not necessarily have the same goals. The organization will strive to impress its pattern upon the individual, and the individual will strive to impress his pattern upon the organization, the two patterns often being at odds. How then are effective leadership and good administration to be achieved? Essentially through the "fusion process": "Effective leadership behavior is 'fusing' the individual and the organization in such a way that both simultaneously obtain optimum self-actualization" (Argyris, 1957, p. 13).

There was not a single formulation of the period that did not insist on the primacy of theory over prescription or that did not try to integrate the personal and impersonal aspects of organization and administration into a unified scheme. That remarkable treatise by Chester Barnard, *The Functions of the Executive*, which had been published in 1938, was rediscovered or really discovered. Barnard conceived of an organization as involving production (or impersonal) forces and humane (or personal) forces. The functions of the executive were to coordinate these forces in order to maximize both effectiveness, the accomplishment of organizational purpose, and efficiency, the satisfaction of individual purpose. So far as I could find, virtually no text in educational administration before the 1950s mentions this work; by the 1960s, hardly a text does not mention it.

3.5 *Concepts and Research in Administration and the Preparation of Educational Administrators*

Similar formulations and syntheses were developing specifically in educational administration. These were set forth at a crucial meeting of the National Conference of Professors of Educational Administration in 1954. Some scholars even point to a single "brief but provocative and influential monograph" by two researchers (Coladarci & Getzels, 1955) significantly entitled *The Uses of Theory in Educational Administration*, having its genesis at the meeting and published in 1955, which "can be identified as the herald of this new era (and) pointed the way for much of the work which has followed" (Jensen & Clark, 1964, p. 59). This seems to attribute too much to a single work or event. But there is no doubt that the meeting and the monograph are benchmarks of the new era. New conceptions of

educational administration and material for preparing school administrators were in the making as the volume *Administrative Behavior in Education* (Campbell & Gregg, 1957), produced under the aegis of the National Conference and published in 1957, was to demonstrate.

I shall cite only two instances of the new theory and research-oriented conceptions that entered into the preparation of educational administrators. One conception, founded in and giving rise to basic research, visualized administrative behavior in the school setting as involving the now familiar dimensions of "Initiating Structure" and "Consideration" (Halpin, 1966). In "Initiating Structure," the administrator lets the staff know what is expected of them and sees to it that they are working up to capacity. In "Consideration," the administrator is friendly, approachable, and gets staff approval on important matters before going ahead. Effective leadership involves the synthesis of Initiating Structure and Consideration. Another conception, again both founded in and giving rise to basic research, visualized the school as a social system that could be divided analytically into a nomothetic or normative dimension and an idiographic or individual dimension; the one referring to the formal structure of roles and expectations in the organization, the other to the informal patterning of personalities and dispositions (Getzels & Guba, 1957). A nomothetic style of administration places emphasis on the production requirements of the system, akin to the scientific management mode; the idiographic style places emphasis on the personal requirements of the individual, akin to the human relations mode. The so-called transactional style synthesizes the two.

3.6 From Prescriptive Solutions to Theoretical Inquiries

This third period of administrative thought in education has often been referred to as the social science era. But the transformation was more profound than only reference to social science would suggest. The essential shift was from a conception—a paradigm—of educational administration as only a vocation to be handed down by word of mouth, as it were, from practitioner to practitioner, to a conception of it also as a domain of study. This conceptual shift moved the focus of effort in the preparation of administrators from a *solution orientation* where experienced administrators gave prescriptive answers to how-to-do-it questions to an *inquiry orientation* where problems were posed and understanding sought regarding the phenomena of administration in their own right. The two successive yearbooks on administration published by the Society for the Study of Education (Griffiths, 1964; Henry, 1946), to which I referred at the beginning, dramatically illustrate the transformation: The 1946 yearbook was largely the work of state school officers and school superintendents; the 1964 yearbook was largely the work of scholars of administration.

There was a concomitant shift in the research during this period. In the first half of the century, research consisted mainly of status studies and collections of opinion. Little research dealt with testing theoretical propositions; virtually none involved the concepts and methods of the behavioral sciences. Later there was a marked reversal; considerable research was directed to testing theoretical propositions, or at least had its source in theoretical formulations and applied the methods and concepts of the behavioral sciences. One need but compare the content of dissertations and journals of the 1930s and 1940s with those of the 1950s and 1960s, especially the articles in the preceding journals in administration with those in the *Journal of Educational Administration* and the *Educational Administration Quarterly*, both of which were founded in this period. The very existence of these journals reflects the changes I have been sketching. As one observer in the 1960s put it, "The current excitement arises from the fact that within recent years educational administration has become a field of study and of development as well as a vocation" (Miller, 1965, p. 545).

We have no way of ascertaining whether the "excitement" in research and theory in administration as a field of study had a beneficial or deleterious effect on the preparation of administrators or the administration of the schools. There are administrators trained during this period who say the research and theoretical views they were required to study had little effect on them (Hills, 1975). There are other administrators trained during the same time and in the same place who say the views had great and highly beneficial effects on them (Faber & Shearron, 1970). But this is not the point here. The point is only that with the conceptual and research shifts came profound shifts in the curriculum and literature used to prepare school administrators.

I have already commented that before 1950 virtually no texts in educational administration mentioned Chester Barnard's work, although it had already been in publication a dozen years. After 1960 virtually no text did not refer to it. The change was not merely in the reference to people but in fundamental point of view. Virtually none of the texts before 1950 referred to systematic theory; virtually none of these after 1960 did not refer to it. Most telling in this respect is to follow the introduction of the new ideas emerging in the theoretical treatises and research reports through successive editions of the same authors' text for preparing school administrators.

I shall illustrate the introduction of the new theoretical and research ideas into the textbooks by referring only to the social system or so-called nomothetic-idiographic model of administration—not because it is the most notable instance of the argument I am making but because I happen to be most intimately acquainted with this model.

Consider a text that went through two editions. In the first edition

there were two references of a sentence each to this formulation, one of which described it as "very limited" (Morphet, Johns, & Reller, 1959). In the second edition, eight years later, the references are quite lengthy, the framework is presented in some detail, and it is described as "a model . . . which has been extremely fertile . . . [and] has already become a classic" (Morphet, Johns, & Reller, 1967, pp. 67-68). Disregarding the merit of any particular model, I suggest rather that there are manifest changes in what students who are preparing to become school administrators are expected to know—changes having their source in basic research and theory.

Consider another instance, this from a text that went through three editions. In the first edition (Knezevich, 1962), the last chapter, which included a description of the social system model, was entitled "Toward the Emergence of a Theoretical Framework for the Study of School Administration." It dealt briefly, almost hesitantly, with theory and its value for administrators, as if the author did not yet know whether to place his bets on the emerging idea. There is nothing hesitant about the second or third editions. By the second edition (Knezevich, 1969), seven years after the first, the chapter originally entitled "Toward the Emergence of a Theoretical Framework . . ." has been entitled "Models in Educational Administration," which ends: "Improvements in and new approaches to school administration can be conceptualized in terms of the administrator-scientist who develops and tests models" (p. 537).

Consider a last example—from a text that has gone through four editions. The foreword to the first edition states quite directly what is needed for the preparation of administrators: "We feel that serious examination of and thought about one's own experiences and those of other people dealing with similar problems is of vital importance to anyone who wishes to gain a real understanding of the administrative process" (Campbell et al., 1958, p. vi). The emphasis is on the experiences of administrators and what new administrators need to know and do. There are sections on the salary schedules for administrators, information about tenure and security, a list of the associations and affiliations (the text names some three dozen organizations) that the new administrator should take into account in choosing membership in a community organization, and so on. There is no mention of theory, models, conceptual schemes, or social science.

In the second edition (Campbell, Corbally, & Ramseyer, 1962), four years later, the sections on salary schedules, tenure and security, organizational affiliation, and such, remain. But a sizable new chapter has been inserted on the nomothetic-idiographic model with a section on the implications of the model for administrative practice. By the fourth edition, nine years later, the foreword states: "The book is built on the conviction that one develops an understanding of educational administration as one

examines *concepts* and relates them to actual experiences" (Campbell et al., 1971, p. xiv). The sections on salary schedules, tenure and security, associations and affiliations have completely disappeared. A substantial section entitled "Theory in Administration" has been added. The nomothetic-idiographic model is presented as a "theory" of administration, and a number of its concepts are utilized to explicate certain administrative problems that face the administrator in practical situations.

Hé, then, is a crucial change in the preparation of school personnel as a function of theoretical research. In the first edition, understanding educational administration came from relating the prospective administrator's experiences to the experiences of already practicing administrators. In succeeding editions, understanding is said to come from relating the prospective administrator's experiences not only, and perhaps not even primarily, to the experiences of others but to theoretical concepts. The index to the first edition carried no references to those responsible for the nomothetic-idiographic model; the second edition carried 7 references; the third, 10; the fourth, 15.

Further, if one assumes, as I believe one may, that Barnard, Simon, and Parsons were the patron saints of organizational and administrative theory at the time, there are some provocative numbers quantifying the change over the years, for those who are convinced only by numbers. The index to the first edition gives no references to Parsons or to Simon, the index to the fourth edition gives five page-references to Parsons and six to Simon. The index to the first edition gives only one reference to Barnard; the index to the second edition gives three; to the third edition, five; to the fourth, seven. In the first edition, Barnard's views are allotted exactly one sentence; by the fourth edition they receive an exposition of several pages.

4. BASIC RESEARCH AND THE UNIVERSITY

Asked of any other institution, What is it about? What does it do? What does it produce? the answer is relatively simple. The work of a factory, a store, a hospital may be described as providing solutions to manifest problems—problems that have already been identified and formulated. The work of a university cannot be described so simply. The work of a university, whether in the classroom, in the laboratory, or in the field, lies distinctively in making explicit that which is problematical—that is, in exploring fundamental enigmas still needing theoretical formulation as problems, to say nothing of solutions (Getzels, 1973).

The university's engagement with the enigmatic rather than the apparent, the ultimate rather than the immediate, the fundamental rather than the pragmatic, renders it liable to the charge that it is not doing anything useful or practical. The charge is of course the same as the one

directed against basic research, which, along with instruction, is the university's essential function. It is especially vulnerable to this charge when the principal criteria of worth become, as in fact they have recently, cost-benefit analyses and the standard of value the practical payoff. One can judge the cost-benefit, the payoff, of a solution that can be applied, of a package that can be sold. But how does one measure the cost-benefit of a problem that is found—a problem, moreover, that may not be solved in one's own lifetime, or if solved may have no apparent application at the moment.

The question may be raised: Need one *find* problems? Is not the world already teeming with dilemmas, at home and abroad, in the arts and in the sciences, in economics and in education? Of course, the world is teeming with dilemmas. But the dilemmas do not present themselves automatically as *problems* capable of resolution or even of sensible contemplation. As Albert Einstein (Einstein & Infeld, 1947) put the issue, "The formulation of a problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old questions from a new angle, requires imagination and marks real advance in science" (p. 95) and, it may be safe to say, in activity of any worth.

4.1 *The Identification of Basic Problems*

It is exactly this activity—the identification of where problems lie and where questions are still needed, and not only the solution of problems already identified—that is at once the distinctive mark of basic research and the work of the university in all areas of inquiry, including education. The products of such work, like the problems themselves, often seem esoteric or, as the saying goes, merely theoretic; and their value is not measurable by the popular criteria of the moment. But they bear fruit in knowledge that illuminates our experience and, in due course, often present applications of the greatest practical utility.

Consider Feynman (1974), the Nobel-prize physicist, describing his work on quarks. He comes to a point where he needs a mathematical formula to move the work forward. He writes, "In fact, the equations (needed) . . . were written down over 20 years ago by Yang and Mills, who saw no application of them but published them because they looked so beautiful and symmetrical" (p. 610). Feynman then adds—and this is my contention regarding the utility of basic research and intellectual products which seem to have no immediate application—"sometimes the truth is discovered *first* and the beauty or 'necessity' of that truth seen only later" (p. 610).

All those who would measure the fruits of inquiry and the scholarly enterprise by immediate application should contemplate and savor this statement: The truth is discovered first and the necessity or utility of that

truth may be seen only later. I know of no better description of what basic research is about—whether it is in mathematics, physics, political science, child development, or education. For if the truth is not discovered first, it will not be there when the practical necessity for it arises later. This of course illuminates the danger of the anti-intellectual standards applied by funding agencies for research in education that hold only to the immediate, measurable, and material consequences in evaluation.

Consider a more practical instance. The invention of the electric dynamo would not have been possible if someone had not posed the prior basic problem (a problem that surely did not trouble anyone in any immediate practical way): What is the nature of electromagnetic induction? Quite typically, many of Faraday's contemporaries looked upon his investigations with skepticism and considered them useless. In fact, Gladstone, then the Chancellor of the Exchequer, interrupted Faraday while he was describing his work to ask impatiently, "But after all, what use is it?" Faraday's response was, "Why, Sir, there is every possibility that you will soon be able to tax it" (Chandrasekhar, 1975, p. 107). In our time, the invention of the polio vaccine would not have been possible if someone had not posed the prior, apparently "useless," problem: What is the nature of the mammalian cell, and can it be maintained outside the living organism.

There are those who argue that such instances may hold in the physical and biological sciences but not in the social sciences, and certainly not in education. One must beware of analogies. Yet what would the present curriculum look like if Thorndike and Woodworth had not posed the problem: What is the nature of transfer of training from one subject to another? Or what would our attitudes toward gifted children be today if Terman had not posed the problem: What is the nature of giftedness and its relation to later achievement? And what would the new elementary science programs have looked like if Piaget's experiments on cognitive conservation had not been performed, since they had no apparent relevance to education when they were undertaken? Or what would the preparation of administrators have been like if Lewin and his colleagues had not posed the problem: What is the nature of different types of leadership and their effect on the dynamics of groups? And in what direction would the preparation of school administrators have proceeded if social systems theory had not been formulated?

4.2 *The Question of Accountability*

The apparent freedom of the basic researcher to pursue his own interests rather than prescribed missions—to work on "discovered problems" rather than "presented problems" (Getzels & Csikszentmihalyi, 1967)—gives the impression that basic research is unplanned, hit or miss, and irresponsible. It appears that any problem will do, and the university

professor or basic researcher is not accountable. Nothing is further from the case. The accountability system in fundamental scholarship is strict and relentless. It is more severe than in any other professional work and surely more severe than in the mission-oriented work so much favored by funding agencies. It is a system integral to the basic research process itself, as much part of the work as the procedures and results. It is a system composed of one's colleagues within the university and in universities far removed; a system of journal editors and anonymous referees, of public review and critical challenge; of recognition, indifference, or ruthless condemnation.

Today there is pressure on the university and more especially on those in education to turn away from exploring fundamental enigmas and turn to delivering technical services, away from formulating theory-oriented problems to providing mission-oriented solutions. The pressure has its source not only in the difficulties facing education but in an unhappy confluence of anti-intellectualism and financial stringency. It takes many forms; above all, the popularity of accounting of results not by their contribution to fundamental knowledge but by their immediate and material consequences.

No one can deny the need for services and mission-oriented solutions. The university and those in it cannot and should not neglect these. Yet there are other institutions and other individuals that can provide services and undertake missions and perhaps do these tasks just as well, if not better. But if those in the university are not permitted to behave distinctively in its mode, if they are not permitted to pursue basic theory-oriented inquiry, who else will ask the fresh questions, explore the fruitful enigmas, the necessity for which (as in the case of Yang and Mills' equations, Terman's studies of giftedness, or Lewin's investigations of group dynamics) may not be immediately apparent, but without which future dilemmas may not be resolved?

5. THEORY, RESEARCH, AND PRACTICE

Having dealt with the relation between basic research and practice, and more especially with the controversy between those who assert such research has little practical bearing on the operation of schools and those who say that such research has had profound practical bearing, there is a second controversy, related to the first, that deserves comment, if only in passing. The issue may be stated most briefly in dichotomous terms, although this runs the danger of oversimplification (Coladarci & Getzels, 1955; Getzels et al., 1968; Getzels, 1975).

On the one hand, there are researchers who begin with the premise that the primary focus of educational inquiry must be on the collection of reliable observations and data. To these researchers the formulation of

theory is as yet premature and diversionary and, in any case, cannot be done without a base in reliable observations and data. On the other hand, there are those who begin with the opposed premise that the primary focus of effort must be the theoretical work of formulating heuristic conceptions to guide the observations to be made and the data to be collected. Collecting data without a base in such conceptions, it is claimed, leads only to "blind" empiricism, and in any case observations cannot be made without recourse to *some* theoretical framework. The obvious and facile solution to the dilemma is that priority not be given to one view over the other; a middle course between the two should be taken, a compromise struck. As will be seen, the solution is more easily advised than achieved.

5.1 *The Question of Theoretical Bias*

The first group points to the inevitable biases of current theories and urges the "deliberate suspension of all implicit and explicit assumptions . . . which might bias our observations" (MacLeod, 1947, p. 194). The attempt must be made to "observe and describe in all its essential characteristics the world of phenomena as it is presented to us." The question to be raised is, "simply, 'What is there?', without regard to Why, Whence, or Wherefore," that is, without regard to any theoretical conception—whether behavioristic, cognitive, interactionist, sociological, or whatever. To proceed otherwise is to invite biases from theory that predetermine what is seen and prejudice the character of the practical solution.

Several categories of theoretical bias may illustrate the point. There is, for example, the "organism centered bias" (MacLeod, 1947) where the determinants of behavior are defined as conditions of the organism or as forces like needs, drives, or dispositions. Why does a pupil work hard? Because of Need: Achievement. Why is he hostile? Because of a disposition for curiosity. There is the "genetic bias" where the reason for a present activity is sought in forces that operated at an earlier period rather than in those operating at the time of the activity. The questions, Why does the pupil work hard? or Why is he hostile? or Why creative? are answered by a "because" pointing back to events in early childhood. There is the "sociological bias" where the human activity becomes defined by a set of conceived institutional structures having reality for the observer but not necessarily for the individual observed. The questions are then answered by a "because" of the presumed institutional structure and process postulated by the sociologist rather than the situation as apprehended and reacted to by the person. To avoid these and other biases, the investigator must proceed by looking at the "world of things-as-they-are in its entirety" and collecting "the facts" without any distorting effect or prior theory.

Those in the second group agree that, of course, one must begin by

collecting "the facts." But they immediately echo Morris Raphael Cohen's (1953) query, "Begin by collecting the facts? Aye, but what facts?" (p. 76). Observations and measurements, no matter how precise, do not of themselves lead to stable, practical knowledge except through the application of some directive concepts, however tentative, that serve as guides as to *what* to observe or measure and *how* to interpret what has been observed and measured. As Larrabee (1945) points out:

The advice to "get the facts" is not a blanket injunction to expose oneself to anything in the way of experience that may come along. It leads to the query: "But what facts?" . . . Relevant facts do not label themselves as relevant. That is an element which must be added by the knower. Unless he is a mere random collector of odds and ends, the seeker of knowledge cannot go through life merely looking *at* things; he must look *for* something; and that means active inquiry with some directing factor in control. (p. 167)

This holds not only for high science but for so-called common sense, the latter also being based on certain assumptions, albeit implicit rather than explicit ones. As Conant (1952) says:

Literally, every step we take in life is determined by a series of interlocking concepts and conceptual schemes. Every goal we formulate for our actions, every decision we make, be it trivial or momentous, involves assumptions about the universe and about human beings. To my mind, any attempt to draw a sharp line between the common-sense ideas and scientific concepts is not only impossible but unwise. (p. 80)

From this point of view, conceptual and theoretical bias in "collecting the facts" is inevitable. The observing, perceiving, thinking organism cannot proceed without an assumptive base. One cannot look at everything, and so one must select what to look at on some grounds. The mind abhors a conceptual vacuum; if one set of assumptions and concepts is suspended, another set, unbeknownst to the observer, will sneak in. Instead of suspending theory, which is impossible (the language one chooses to use is itself "theoretical"), one should proceed within the framework of consciously held theory. In Neal Miller's (1959) blunt words:

Pure empiricism is a delusion. A theory-like process is inevitably involved in drawing boundaries around certain parts of the flux of experience to define observable events and in the selection of events to be observed. Since multitudinous events could be observed and an enormous number of relationships could be determined among all these events, gathering all the facts with no bias from theory is utterly impossible. Scientists are forced to make a drastic selection, either unconsciously on the basis of perceptual habits and the folklore and linguistic categories of the culture, or consciously on the basis of explicitly formulated theory. (p. 200)

5.2 *Consequences of Avoiding Theory*

This juxtaposition of the two views in extreme form probably draws the line between them too sharply, but that the issue exists is certain.

There are a number of untoward practical and research consequences in education that follow from inattention to the issue, and more especially from the avoidance of theory (Coladarci & Getzels, 1955).

One consequence is *factualism*. The attempts to make education as "scientific" as possible have been directed merely to making it as "factual" as possible. (Smith, 1950). Thus, both research and practice have been pushed more and more into empiricism without the support of theoretical and explanatory concepts. As a result, studies are often no more than journalistic descriptions or collections of interesting, and even precise, but disconnected and irrelevant data that are heralded as facts. Such data or "facts" do not themselves carry explanations of their import or directions for their application.

A second consequence of avoiding theory is *excessive reliance on the authority of experts, principles, and techniques*. Consider, for example, the frequency with which educational research and evaluation incorporate as a validating criterion "the opinion of experts in the field," or consider the extent to which principles and techniques of educational practice are defended by the assertions that they have been recommended by this or that expert. It is not that experts may not be used judiciously as consultants, but it must be understood that there is a difference between relying on the magisterial assertions provided by experts and testing the hypothetical statements engendered by theory.

A third consequence is *lack of adequate professional language*. An explicit theory requires the development of rigorously defined concepts and the use of an unambiguous language for communication. Without theory, any old terms and any old notions may be bandied about as if they had been explicitly defined. It is for this reason that the language of education is often accused of being "jargon" or "gobbledygook." It is not a defined specialized terminology in a field that makes for jargon—as witness physics ("quark") or psychoanalysis ("Id")—but its absence. Consider, for example, the term "need." In psychology it has a specific meaning referring to an explicitly defined concept in a particular theory of personality. One may disagree with the theory or with the usefulness of the concept but no one would accuse the language of being gobbledygook. In education, the same term used in the familiar "meeting the *needs* of children" is left undefined and seems to have as many meanings as there are people who use it—or no meaning. What *does* "meeting the needs of children" mean? Does it refer to personalistic needs like "Need: Achievement" or "Need: Affiliation" as theoretically defined? Or does it refer to what children "lack," "want," or what they "require" whether they want it or not? "Meeting the needs of children" as used in education without theoretical context to give the phrase meaning is justly accused of being jargon and a *mère* cliché.

A final concomitant of avoiding theory is *emotional identification with*

personal views. When the dialogue of competing theories is foregone, one is in danger of becoming emotionally and blindly wedded to one's own observations. Events seem really to be what one perceives them to be. This precludes the possibility of looking at one's ideas as tentative, open to criticism, or in need of testing. Those who eschew the theoretical tend to hold their beliefs not as hypotheses but as cherished possessions and are unable to participate in the evaluation and reformulation of the assumptions on which their observations and decisions are based. As Gordon Allport (1955) put it, "Our [strongest] censure should be reserved for those who close all doors but one. The surest way to lose truth is to pretend that one already wholly possesses it" (p. 17).

5.3 *Objections to Utilizing Theory*

These are among the pernicious concomitants of avoiding theory. But there are also a number of objections to utilizing theory. Among those more frequently mentioned are the following: Theory restricts one's freedom of judgment and leads to "push-button" decisions; since so many theorists are available and they are so short-lived, there is little point in giving them serious consideration; the imposition of a theoretical structure prohibits the unbiased observation of phenomena; there is the danger of becoming irrevocably identified with a single theory (Coladarci & Getzels, 1955; Getzels et al., 1968).

The objection that consciousness of theory will restrict the practitioner's (or researcher's) freedom of judgment and lead to push-button decisions may be dismissed as untenable. The awareness of theories, to use an analogy, is like the knowledge of maps. The understanding of possibilities provided by relevant theories, like the intelligence of alternate routes supplied by maps, allows the practitioner the greatest freedom to make informed choices and apply his judgment, experience, and ingenuity. It is the application of a prescription without a framework, like the pursuit of an itinerary without a context, that limits choice and inventiveness and removes the "art" from practice.

The objection that there is little point in giving theories serious consideration since there are so many of them and they are so short-lived is also untenable. As Neal Miller (1951) writes, "The value of a theory is not measured by the length of its life. A theory can be wrong and still lead to progress" (p. 88). The history of physics provides excellent examples of theories that changed rapidly—that is, to use Miller's term, were "wrong"—but were useful. No sooner was a theory explicitly formulated than new observations prompted by it led to its revision, and a new theory was formulated on the basis of the experience with the old one. But surely the old theory was not a waste of effort; it gave rise to hitherto unthought of directions for observation and brought to light otherwise unavailable possibilities for action. In this sense, explicit theory—even wrong

theory—is more fruitful than implicit theory or the belief in no theory (“just the facts”) to which education is so prone.

The third and fourth objections—that the imposition of a theoretical structure may bias the observations of phenomena, and that there is a danger of becoming emotionally identified with a given theory—are more serious. Theoretical assumptions may in fact prejudice our observations. As we have seen, an organism-centered theory of behavior will point to the source of human action in forces embedded in the organism, excluding or obscuring other possibilities; a genetic theory will point to the source for the same action in forces that operated in previous times, again excluding or obscuring other possibilities; a sociological theory will point to the source in the institutional structure of society, also excluding or obscuring other possibilities; and so on. Originators and proponents of these conceptions may become so emotionally identified with their particular view that when events do not conform to their formulation, instead of altering their theory, they alter their view of the phenomena to conform to the formulation they have invented.

5.4 *Theoretical Safeguards*

These objections are well taken and serious. But there are certain safeguards built into the theory process itself as well as certain precautions that the user of theory can take. The assumptions, concepts, and methods of the various theories are out in the open, whereas those of the experts, authorities, and observers professing to see the universe “as it really is” are not. The theoretical assumptions, concepts, and methods can be examined for possible biasing and limiting effects. We know what to expect from behavioristic, cognitive, or sociological theory, and if the partial views and distorting effects of each are unavoidable, they are recognizable and can be taken into account. There is also a difference between the closed-end assertive statements of authority and the open-end hypothetical statements of theory. Competing authorities produce debate, the arbiter often being only a reference to other authority. Competing theories engender dialogue inviting inquiry—indeed, basic or theory-oriented research is precisely such inquiry—and the ultimate arbiter is the world of events. A conceptual formulation, however attractive, that distorts the world of events will be superseded, as the rise and fall of theories in all fields of study will attest, just as surely as a map that misrepresents the terrain is superseded.

None of this is to minimize the pitfalls. Nor is it to suggest that the insights of theory, any more than the results of research, can be converted automatically into general directions for practice or immediate rules of educational art (Dewey, 1929). The insights of theory, like the results of research, must inevitably be particularized by the practitioner in the light of his own circumstance and experience. Yet as Coladarci and Getzels (1955) observed, “The educator who behaves on a hit-or-miss basis, one

whose professional arsenal consists merely of pat techniques for specific situations, is operating in intellectual low gear. . . . A clarified and well-thought-out theory, no matter how provisional, is a frame-of-reference that creates some order out of what otherwise might appear to be a disorganized situation (inviting only) trial-and-error behavior" (pp. 7-8). John Dewey (1929) put the issue as follows:

Theory is in the end . . . the most practical of all things, because the widening of the range of attention beyond nearby purpose and desire eventually results in the creation of farther and wider-reaching purposes, and enables us to make use of a much wider and deeper range of conditions and means than were expressed in the observation of primitive practical purposes. (p. 17)

6. SUMMARY

Let me summarize what I have tried to do and add a word of qualification lest I be misunderstood.

I began by remarking on the widely held skepticism regarding the influence of research and more particularly of basic research on educational practice. Even at its best, it is said, basic research produces only "tiny bits" of knowledge without substantial relevance to the actual operation of schools. The consequences of this have been the sacrifice of support for research and especially basic research in favor of applied and mission-oriented activities, which presumably would have immediate bearing on the operation of schools.

My argument was to the contrary, that is, basic or theory-oriented research does have a profound bearing on the operation of schools.

First, by reference to specific studies, I showed that fundamental research has powerful effects on the school; indeed, the nature of certain components of the curriculum, the attitudes of teachers toward students, and the forms of classroom management, among other aspects of the operation of schools, are a function of the findings of theoretical inquiry and cannot be fully understood apart from them. Among the illustrations were the effects of basic studies in transfer of training on the role of Latin in the curriculum; the studies of giftedness on our attitudes toward precocious pupils; the studies of autocratic, democratic, and laissez-faire leadership on the management of classrooms; and the studies of cognitive development on innovations in elementary science materials.

Second, far from contributing only tiny bits of knowledge, as charged, basic or theory-oriented research contributes to the broad conceptions, the paradigms, of the human being and his universe and thus provides fundamental contexts for education. I illustrated this by showing how certain changes in the aspirations and forms of the classroom since the turn of the century were concomitant with certain changes in the conceptions of the child as learner, changes that were a result of theoretical inquiry. The successive changes in the conformation of the ideal learning

environment from the "straight-row rectangular" classroom to the "square," "circular," and "open" classroom cannot be fully understood apart from the influence of the changes in the conception of the learner from an "empty" organism to an "active," "social," and "stimulus-seeking" organism.

Third, not only does theoretical research enter into changes in the curriculum and the conformation of the classroom, but it enters also into the preparation of school personnel. Was there any textbook for the preparation of teachers that did not at one time include the laws of readiness, exercise, and effect, or later the principles of group dynamics and classroom climate, and still more recently the stages of cognitive growth? As a more concrete case instance, I showed that as research and theories of organizational behavior in the school as a social system evolved, they were methodically introduced into the curricula and textbooks used to prepare school administrators, and a formulation that was dismissed as irrelevant or disregarded entirely in the first edition of a textbook was made central to what a school administrator should understand and presumably act on in subsequent editions.

Fourth, I addressed the role of the university in the research enterprise—a role that is being increasingly and grievously enfeebled by mistaken beliefs regarding the separation of basic research from practical application and the accounting of results of inquiry not by their contribution to fundamental knowledge but by their immediate material consequence. The work of the university and its scholarship lies uniquely, as does that of no other institution, in making explicit that which is problematic, in exploring enigmas lacking formulation as problems, to say nothing of solutions. This identification of problems and of questions that must be asked (not only the solution of questions already asked) requires, as Einstein put it, imagination and marks the real advances of science and, one may add, of all activities of any worth. This, then, may be the paradox: Research may have the greatest effects on education (as it has had elsewhere) not so much where it attempts to alter an element of practice directly (in the mode of the packager), as indirectly (in the mode of the University) where it raises new questions and contributes to transformations in the general paradigms of the human being and his condition.

Finally, I commented briefly on two competing views held by educators regarding the relations of theory, research, and practice. On the one hand, there are those who would eschew theory since it is said to bias observation of the world as it is. On the other, there are those who insist that pure empiricism is a delusion and eschewing theory is not only undesirable but impossible. Each view has strategic consequences. Among the pernicious concomitants of eschewing theory and theory-oriented or basic research are factualism, reliance on experts, adherence to pat

techniques, lack of a professional language, and emotional identification with personal opinion.

My contention is not that basic research inevitably has practical uses, or that the practical uses of the basic research to which I referred have invariably been salutary—it should be noted too that theoretic knowledge is unlikely to become practically useful except through the judgment, skill, and experience of the practitioner who applies the general concepts to the particular circumstance. Nor am I contending that development of mission-oriented material and curricula is not needed, or that research deriving from immediate classroom problems—say, whether to use one procedure as against another in teaching a particular subject—is without effect. Such research and development are manifestly of value; they require no defense. But in light of the increasing number of assertions that basic or, as it is invidiously called, “merely theoretical” research is of little value for education, coupled with the increasing sacrifice of support for basic research for presumably more practical activities, I contend it is necessary to consider in specific terms the view that even for the curriculum, structure, and administration of schools there may ultimately be, to paraphrase Dewey, nothing as practical as research that is “merely theoretical.”

REFERENCES

- Allport, G. W. *Becoming*. New Haven, Conn.: Yale University Press, 1955.
- Argyris, C. *Personality and organization*. New York: Harper & Row, 1957.
- Bakke, E. W. *Organization and the individual*. New Haven, Conn.: Yale University, Labor and Management Center, 1952.
- Barnard, C. *The functions of the executive*. Cambridge, Mass.: Harvard University Press, 1938.
- Barth, R. S. *Open education and the American school*. New York: Agathon Press, 1972.
- Berlyne, D. E. Curiosity and exploration. *Science*, 1966, **153**, 25–33.
- Bexton, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in sensory environment. *Canadian Journal of Psychology*, 1954, **8**, 70–76.
- Burks, B. S., Jensen, D. W., & Terman, L. M. *Genetic studies of genius: The promise of youth* (Vol. 3). Stanford, Calif.: Stanford University Press, 1930.
- Campbell, R. F., Bridges, E. M., Corbally, J. E., Nystrand, R. O., & Ramseyer, J. A. *Introduction to educational administration* (4th ed.). Boston: Allyn & Bacon, 1971.
- Campbell, R. F., Corbally, J. E., & Ramseyer, J. A. *Introduction to educational administration*. Boston: Allyn & Bacon, 1958.
- Campbell, R. F., Corbally, J. E., & Ramseyer, J. A. *Educational administration*. Boston: Allyn & Bacon, 1962.
- Campbell, R. F., & Gregg, R. T. *Administrative behavior in education*. New York: Harper & Row, 1957.
- Carroll, J. B. *Educational psychology: Achievements and prospects*. Paper presented at the meeting of the National Society of College Teachers of Education, Chicago, Feb. 14, 1963.

- Chandrasekhar, S. Shakespeare, Newton, and Beethoven, or patterns of creativity. *University of Chicago Record*, 1975, **9**, 91-108.
- Cohen, M. R. *Reason and nature* (2nd ed.). New York: Free Press, 1953.
- Coladarci, A. P., & Getzels, J. W. *The use of theory in educational administration*. Stanford, Calif.: Stanford University, School of Education, 1955.
- Colvin, S. S. *The learning process*. New York: Macmillan, 1913.
- Conant, J. B. *Modern science and modern man*. New York: Columbia University Press, 1952.
- Cremin, L. *The transformation of the school*. New York: Knopf, 1961.
- Cronbach, L. J., & Suppes, P. (Eds.). *Research for tomorrow's schools*. (Report of the Committee of Educational Research of the National Academy of Education.) New York: Macmillan, 1969.
- Cubberley, E. P. *Public school administration*. Boston: Houghton Mifflin, 1916.
- Dewey, J. *Sources of a science of education*. New York: Liveright, 1929.
- Educational Facilities Laboratories. *The cost of the schoolhouse*. New York: Author, 1960.
- Edwards, A. L. Political frames of reference as a factor influencing recognition. *Journal of Abnormal and Social Psychology*, 1941, **36**, 34-50.
- Einstein, A., & Infeld, L. *The evolution of physics*. Cambridge, England: Cambridge University Press, 1947.
- Faber, C. F., & Shearron, G. F. *Elementary school administration*. New York: Holt, Rinehart & Winston, 1970.
- Feynman, R. P. Structure of the proton. *Science*, 1974, **183**, 601-610.
- Getzels, J. W. Creative thinking, problem-solving, and instruction. In E. R. Hilgard (Ed.), *Theories of learning and instruction*. (Sixty-third Yearbook of the National Society for the Study of Education, Part 1.) Chicago: University of Chicago Press, 1964.
- Getzels, J. W. The nature and nurture of the gifted child. In R. E. Cooke & S. Levin (Eds.), *The biologic basis of pediatric practice* (Vol. 2). New York: McGraw-Hill, 1968.
- Getzels, J. W. A social psychology of education. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 5). Reading, Mass.: Addison-Wesley, 1969.
- Getzels, J. W. On the transformation of American values: A decade after Port Huron. *School Review*, 1972, **80**, 505-519.
- Getzels, J. W. Problem finding. *University of Chicago Record*, 1973, **7**, 281-283.
- Getzels, J. W. Creativity: Prospects and issues. In I. A. Taylor & J. W. Getzels (Eds.), *Perspectives in creativity*. Chicago: Aldine, 1975.
- Getzels, J. W., & Csikszentmihalyi, M. Scientific creativity. *Science Journal*, 1967, **3**, 80-84.
- Getzels, J. W., & Dillon, J. T. The nature of giftedness and the education of the gifted. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand-McNally, 1973.
- Getzels, J. W., & Guba, E. G. Social behavior and the administrative process. *School Review*, 1957, **65**, 423-441.
- Getzels, J. W., Lipham, J. M., & Campbell, R. F. *Educational administration as a social process*. New York: Harper & Row, 1968.
- Griffiths, D. E. (Ed.). *Behavioral science and educational administration*. (Sixty-third Yearbook of the National Society for the Study of Education, Part 2.) Chicago: University of Chicago Press, 1964.
- Guba, E. G., & Clark, D. L. The configurational perspective: A new view of

- educational knowledge production and utilization. *Educational Research*, 1975, **4**, 6-9.
- Halpin, A. W. *Theory and research in administration*. New York: Macmillan, 1966.
- Harlow, H. F. Motivational forces underlying behavior. In Kentucky Symposium, *Learning theory, personality theory and clinical research*. New York: Wiley, 1954.
- Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.
- Hebb, D. O., & Thompson, W. R. The significance of animal studies. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1). Reading, Mass.: Addison-Wesley, 1954.
- Henry, N. B. (Ed.). *Changing conceptions in educational administration*. (Fifty-fifth Yearbook of the National Society for the Study of Education, Part 2.) Chicago: University of Chicago Press, 1946.
- Hilgard, E. R. *Theories of learning*. New York: Appleton-Century-Crofts, 1948.
- Hills, J. The preparation of administrators: Some observations from the "firing line." *Educational Administration Quarterly*, 1975, **11**, 1-20.
- Horne, H. H. The practical influence of the new views of formal discipline. *Education*, 1909, **29**, 614-623.
- Hugh, D. D. Formal education from the standpoint of physiological psychology. *Pedagogical Seminary*, 1898, **5**, 599-605.
- Hunt, J. McV. Experience and development of motivation: Some reinterpretations. *Child Development*, 1960, **31**, 489-504.
- Jensen, T. J., & Clark, D. L. *Educational administration*. New York: Center for Applied Research in Administration, 1964.
- Jung, C. G. The gifted child. In H. Read, M. Fordham, & G. Adler (Eds.), *The collected works of C. G. Jung* (Vol. 17). New York: Pantheon, 1954.
- Karplus, R. The science curriculum improvement study—Report to the Piaget conference. *Journal of Research in Science Teaching*, 1964, **2**, 236-240.
- Keynes, J. M. *The general theory of employment, interest, and money*. New York: Harcourt, Brace & World, 1964.
- Knezevich, S. J. *Administration of public administration*. New York: Harper & Row, 1962.
- Knezevich, S. J. *Administration of public administration* (2nd ed.). New York: Harper & Row, 1969.
- Kuhn, T. S. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.
- Lange-Eichbaum, W. *The problem of genius* (E. Paul & C. Paul, trans.). New York: Macmillan, 1932.
- Larrabee, H. A. *Reliable knowledge*. Boston: Houghton Mifflin, 1945.
- Levine, J. M., & Murphy, G. The learning and forgetting of controversial material. *Journal of Abnormal and Social Psychology*, 1943, **38**, 505-517.
- Lewin, K., Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created "social climates." *Journal of Social Psychology*, 1939, **10**, 271-299.
- Lombroso, C. *The man of genius*. London: Scott, 1891.
- MacLeod, R. B. The phenomenological approach to social psychology. *Psychological Review*, 1947, **54**, 193-210.
- McCaul, R. L. Personal communication, June 23, 1973.
- Metcalf, H. C., & Urwick, L. *The collected papers of Mary Parker Follett*. New York: Harper & Row, 1940.
- Miller, N. E. Comments on theoretical models: Illustrated by the development of a theory of conflict behavior. *Journal of Personality*, 1951, **20**, 82-100.

- Miller, N. E. Liberalization in basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of science* (Vol. 2). New York: McGraw-Hill, 1959.
- Miller, V. *The public administration of American school systems*. New York: Macmillan, 1965.
- Morphet, E. L., Johns, R. L., & Reller, T. L. *Educational administration: Concepts, practices, and issues*. Englewood Cliffs, N.J.: Prentice-Hall, 1959.
- Morphet, E. L., Johns, R. L., & Reller, T. L. *Educational administration: Concepts, practices, and issues* (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1967.
- Nisbet, J. F. *The insanity of genius and the general inequality of human faculty physiologically considered*. London: Kegan Paul, 1891.
- Piaget, J. *The origins of intelligence in children*. New York: International Universities Press, 1952.
- Piaget, J. *The construction of reality in the child*. New York: Ballantine, 1954.
- Postman, L., Bruner, J. S., & McGinnis, E. Personal values as selective factors in perception. *Journal of Abnormal and Social Psychology*, 1948, **43**, 142-154.
- R & D Perspectives*. Eugene, Ore.: Center for the Advanced Study of Educational Administration and the ERIC Clearinghouse on Educational Administration, 1969.
- Roethlisberger, F. J., & Dickson, W. J. *Management and the worker*. Cambridge, Mass.: Harvard University Press, 1939.
- Schachtel, E. G. *Metamorphosis*. New York: Basic Books, 1959.
- Segal, E. M., & Lockman, R. Complex behavior or higher mental processes: Is there a paradigm shift? *American Psychologist*, 1972, **27**, 4f-55.
- Sherman, J. G. The theory behind PSI. In J. G. Sherman (Ed.), *PSI personalized system of instruction: 41 germinal papers*. Menlo Park, Calif.: Benjamin, 1974.
- Simon, H. A. *Administrative behavior*. New York: Macmillan, 1945.
- Smith, B. O. Science of education. In W. S. Monroe (Ed.), *Encyclopedia of educational research* (Rev. ed.). New York: Macmillan, 1950.
- Sommer, R. *Personal space: The behavioral basis of design*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.
- Start, K. B. Mirror, mirror . . . reality for the researcher. *American Educational Research Journal*, 1975, **12**, 323-336.
- Stiles, L. J. Developing a research component for education. *Journal of Educational Research*, 1972, **65**, 197-203.
- Stivers, P. NIE: Learning about Congress the hard way. *Educational Researcher*, 1973, **2**, 8-9.
- Taylor, F. W. *Shop management*. New York: Harper & Row, 1911.
- Terman, L. M. *Genetic studies of genius. Mental and physical traits of a thousand gifted children* (Vol. 1). Stanford, Calif.: Stanford University Press, 1925.
- Terman, L. M. The discovery and encouragement of exceptional talent. *American Psychologist*, 1954, **9**, 221-230.
- Terman, L. M., & Oden, M. H. *Genetic studies of genius. The gifted child grows up: Twenty-five years' follow-up of a superior group* (Vol. 4). Stanford, Calif.: Stanford University Press, 1947.
- Terman, L. M., & Oden, M. H. *Genetic studies of genius. The gifted group at mid-life: Thirty-five years' follow-up of the superior child* (Vol. 5). Stanford, Calif.: Stanford University Press, 1959.
- Thorndike, E. L. *Educational psychology* (Vol. 2). New York: Teachers College, 1913.
- Thorndike, E. L., & Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions (Part 1). *Psychological*

- Review*, 1901, 8, 247-261. (a)
- Thorndike, E. L., & Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions (Part 2). *Psychological Review*, 1901, 8, 384-395. (b)
- Thorndike, E. L., & Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions (Part 3). *Psychological Review*, 1901, 8, 553-564. (c)
- Wallen, N. E., & Travers, R. M. W. Analysis and investigation of teaching methods. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand-McNally, 1963.
- White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, 66, 297-333.
- Wise, A. E. The taming of the National Institute of Education: A personal view. *Phi Delta Kappan*, 1976, 58, 62-65.
- Yauch, W. A. *Improving human relations in school administration*. New York: Harper & Row, 1949.

Chapter 8

EDUCATIONAL RESEARCH AND EDUCATIONAL REFORM: A CASE STUDY OF SWEDEN

by

TORSTEN IIUSÉN

*Institute for the Study
of International Problems in Education
University of Stockholm*

1. THE FRAMEWORK FOR POLICYMAKING

THE extensive inquiries as well as planning activities preceding national legislation in education and affecting the entire system of education in the country are to a large extent carried out by ad hoc committees, so-called Royal Commissions, which constitute a unique element in Swedish policymaking. When fundamental public policy issues emerge, such as basic principles of income taxation, provisions of medical services, structure of the school system, or requirements for university admission, pressures begin to build up for the setting up of a governmental committee that is expected to take a deep look into the problem area. The responsible cabinet member brings the matter up with his colleagues and eventually gets the approval of the cabinet to appoint a Royal Commission to conduct a comprehensive inquiry into the matter. The cabinet also agrees on rather detailed terms of reference for the Commission, which in some cases even prescribes the policy according to which the recommendations are to be made. But in most cases the Commissions are free to conduct their inquiries completely unbiased. In selecting the members of a Commission a wide consultation is conducted with the interested parties. The Government tries to see to it that the various interest groups are represented in a reasonably balanced way. Members of the Parliament from various political parties represent a large group of Commission members. Furthermore, central federations of trade unions as well as other national

associations that are potential pressure groups are represented. A Commission operates a secretariat with access to the expertise of specialists whom they, with approval of the Minister, can call upon to serve on the Commission. Since the 1940s, these governmental committees have increasingly commissioned research focused on central problems within the domain under scrutiny. Thus, several key Commissions, whose recommendations since 1945 have been instrumental in the reform of the Swedish educational system, have initiated some of the most important research projects in education conducted in the country.

The very modus operandi of a Commission has often been conducive to a rather close liaison between research workers on the one hand and policymakers on the other. The following example, which the author himself had an opportunity to experience, may serve as an illustration. The 1957 Commission whose work provided the basis for the 1962 Education Act, which made provision for the introduction of comprehensive education in the entire country after a 10-year period of pilot programs, decided to sponsor an extensive survey of the school system of the City of Stockholm. The overall aim of the 9-year longitudinal survey, with measures of student progress and attitudes taken at the end of each school year, was to assess the outcomes of the introduction of the comprehensive school structure. The City Council had in 1955 decided to introduce the new structure on a pilot basis in the southern half of the city, while the northern half kept the traditional parallel structure at the lower secondary level, which meant that the upper grades of the elementary school ran parallel with the lower grades of the academic secondary school to which transfer could take place after 4 years of elementary school. The design of the survey as well as its progress was discussed regularly over a 3-year period in the full Commission, which was presided over by the Minister of Education. Therefore, without delving too deep into technicalities, policymakers on the Commission had ample opportunity to familiarize themselves with research strategies and methodologies employed by the educational research workers as well as with their theoretical and interpretative frameworks.

The reports as well as the supporting research evidence of a Commission are immediately after publication sent out for evaluation to all kinds of interested public and private bodies. Thus, when the work of drafting a bill is started in the Ministry, it has at its disposal not only a report, which in most cases is supported by a solid majority within the Commission, but also sometimes over a hundred written submissions from the most varied bodies and parties interested in the matter. Particularly in a democracy cherishing consensus this is a procedure that is conducive to it.

Since the members of a Commission who are politicians often have been chosen because they are members of the relevant Committee in the Parliament, they can bring the expertise they have acquired after years of

work on the Commission to bear on the debates about the bill that occur both in the Parliament Committee and when the bill reaches the floor. The Parliament Committee can also call upon experts, for example, educational researchers, to witness. This happened to my knowledge the first time in 1950, when the first Comprehensive Education Bill was under debate and when a 10-year pilot program was contemplated (Husén, 1962). A key issue that then was foremost in the minds of both the proponents and the opponents of the reform bill was the proportion of an age cohort that could profit from academic secondary education and the actual size of the "reserve of ability."

Thus, as far as education is concerned, the liaison between policymakers and researchers has for a couple of decades been a rather close one. The contacts that once were established in the ad hoc Commissions have later to some extent become institutionalized by the setting up of various bodies that serve in an advisory capacity within the central agencies in charge of servicing the educational system at various levels. For example, the National Board of Education and its Bureau of Research and Development have an Advisory Council consisting of representatives of various "consumers" groups, such as members of Parliament and teacher associations.

We shall now proceed to a brief description of how lines of communication were opened up and reinforced and how the relations between research workers and policymakers developed after 1940.

The 1940 School Commission approached the holders of the chairs of education and psychology at the Swedish universities and asked for their evaluation of "the present conception of psychological research on mental development in children and adolescents." The Commission consisted mainly of school and university teachers, but no members represented educational research, and there were no politicians. Much time was spent on the problem of the proper age or grade level for primary and secondary education, on how to deal with the dualism between the two systems. The Commission split into two groups, one in favor of a common 6-year elementary school, upon which all types of further schooling should build, and one in favor of so-called double transfer, that is, one connection after the fourth and another after the sixth year in the elementary school. The issue boiled down to the problem of the "effectiveness" of the four-plus-five and the six-plus-four systems, respectively. The Commission made extensive evaluations of the two systems in terms of actual number of school years to graduation, number of grade repeaters, social background of students, school marks, and age of graduation (Husén, 1962).

The issue which for decades has been a dominant preoccupation of Swedish as well as other European secondary-school teachers, who were used to teaching an intellectually and socially select group, could schematically be stated like this: At what age would it be necessary for pedagogical

reasons to differentiate between the academic and the practical students? With an American vocabulary the problem might be phrased as that of the age at which homogeneous grouping should be introduced or at which the college-bound students should be separated from the rest.

The 1940 Commission wanted the chair holders to give the evidence of psychological research pertaining to this problem. Except for the rather limited evaluation inquiries mentioned above, no research was conducted by the Commission. The professors relied heavily upon the general literature on developmental research and for the rest fell back upon "the depth of their consciousness."

The appointment of the 1946 School Commission, which consisted mainly of members of Parliament, implied a turning point in various respects. Apart from the fact that the Commission politically, and not least in school policy, held rather radical views, it also was the first to launch an extensive research to elucidate the development of "practical" and "academic" aptitudes and their interrelations at various levels.

The Commission expressed a strong belief in educational research in its main report that was turned in to the government in 1948: "The findings of educational and psychological research have to be allotted a decisive role as concerns problems which have been subjected to rigid investigations."

The research program was commissioned to Professor John Elmgren at the University of Gothenburg, who called it a study of "ability structure and factorial maturity in Swedish school children." His investigation (Elmgren, 1952), which was reported by the 1946 School Commission in a separate monograph, was based on cross-sectional surveys of school children from 7 to 16 years of age. Elmgren found, like his predecessors, that so-called practical and theoretical aptitudes were positively correlated. Furthermore, according to the factor analyses, "practical" aptitude did not seem to be "as simple and unitary in its structure as general intelligence." The Commission drew two conclusions from this in its report in 1948. The earlier conception of theoretical and practical aptitudes as being in a compensatory way related to each other was not valid. That finding certainly did not support a parallel school structure; an early differentiation by the transfer of the "academics" to a selective secondary school was not justified.

The interesting feature so far of the expectations put by the policymakers upon the researchers was that they should give generalized answers pertaining to school structure and organizational differentiation. They were expected to specify when children's ability structure had reached the level of differentiation that made a separation into different types of schools justified from an educational point of view. They were supposed to be able to state the appropriate age, in general, for introducing instruction in foreign languages, for example.

Even though the interaction between school and society was appreciated, this process was thought of in a rather static way. Factors such as social-class differences in the choice of educational and vocational careers were hardly yet realized by the policymakers, and researchers were just beginning to study these problems. Irrespective of whether those involved had a conservative or radical outlook upon school organization, they thought that there was a fixed number of students able to profit from academic instruction at the pre-university and university level. The dynamic society was still not recognized as an important factor with its insatiable demand for education, nor was education conceived of as an investment in human capital.

The major argument in favor of a reform of secondary schools and higher education was that of providing equality of opportunity. The essence of "democratizing" the schools was that everybody, irrespective of social, economic, or geographical factors, should be on equal footing to realize his potentialities. This initiated two new lines of inquiry, the study of the reserves of ability and the social factors which, in a selective educational system, acted as barriers against the realization of the principle of equality of opportunity.

An attempt to come to grips with the "reserve of talent" problem was made by Husén (1948), utilizing data on conscripts; at that time virtually every young man by the age of 20 had to undergo a psychological and medical examination before he was considered for military service. Since conscripts' records included mental tests, school history and marks, as well as social background, estimates were made similar to those later made by Wolffe (1954) for the United States. These estimates were quoted by the Minister of Education when he introduced the 1950 Education Bill before Parliament.

Härnqvist (1958) pursued the problem under the auspices of the Royal Commission on Universities. He followed one quarter of the 1945 fourth-grade students up to 1956 or to the age of 20. Relying upon these estimates of the untapped reserve of talent, strong support was given to more generous policies of student grants and loans for university study (provisions that were extensively widened in a 1964 Student Aid Act).

Investigations of the effects of "social handicaps" were launched in the mid-1940s, some of it being part of projects taken up by the institutions of education and sociology, part of it sponsored by ad hoc commissions. Boalt (1947) investigated the role of social handicaps in the school careers of a complete age group of youngsters in Stockholm, following them up from the age of 10 to the age of 20. Orring (1959) studied grade-repeating and dropout in the academic secondary school. Husén (1966) reworked the data and related failure to social background. Studies of this type threw doubt on the alleged superiority of a selective and competitive school system in a society with an expanding economy and a strong need

for mass education at the upper secondary level. When the relative merits of a selective and a comprehensive system are weighed against each other, one also has to consider the price that is paid for the quality of the minority that reaches the top of the ladder. Social bias in the selection and screening procedures in the academic secondary school was obvious.

These investigations played an important role when the 1957 School Commission prepared its report and when the 1962 Education Bill introducing the common 9-year school all over the country was drafted. Similar considerations were also behind the 1964 Act concerning *gymnasium* education.

Shortly after the government had appointed the 1957 School Commission which, according to its terms of reference, had to consider the objectives and thereby the curriculum content of the compulsory school, the Industrial Council for Social and Economic Studies, a private organization supported by the Swedish Employers Association, proposed a research program to be sponsored jointly with the Commission, that is, by the government and the Council. According to the initial plans, the curriculum content of mathematics and Swedish was to be the object of the study. But at the request of the Commission the scope was widened, and three other subjects—physics, chemistry, and civics—were included. The research program was commissioned to the Institute of Educational Research at the School of Education in Stockholm in partnership with the Council for Social and Economic Studies and the 1957 School Commission. Teachers in the public elementary or secondary schools, as well as teachers in the vocational schools and in the *gymnasias*, were asked to rate topic by topic what requirements they put on the students they received from schools providing basic education. They were also asked to rate the shortcomings. A series of private industrial and commercial enterprises allowed the researchers to give an extensive achievement-test battery to the employees during working time. The employees also completed a questionnaire indicating the frequency of use of the various skills and topics of knowledge under investigation.

An advisory council consisting of representatives of teacher associations, management, and labor unions was set up for the project. The representatives agreed to send letters of recommendation to their members. In order to make appropriate preparations for the field study which could be expected to cause some concern (among other reasons because the companies had to pay for a full day's work for thousands of employees who were at the researchers' disposal for 5 hours of achievement testing), a special committee was set up representing the various labor-market organizations. The teacher associations were also represented on the various expert committees which advised on the construction of questionnaires and tests.

The main issue in Swedish school policy from 1940 to 1962 was the

problem of school differentiation (Husén, 1962). It could briefly be stated like this: At what age should the students who are considered to be academically talented be separated from their nonintellectual classmates? Should they transfer to another program in a separate school, to another program taught within the same school, or simply to separate classes within the same program in the same school? The theory held by most Swedish secondary-school teachers was that both categories of students would gain from an early separation. The theoretically oriented students would not be hampered by their slow-learning classmates, and the latter would be saved the feeling of inferiority that would be caused by the constant confrontation with their bright peers.

This problem was investigated by the 1957 School Commission which sponsored an extensive follow-up project which has been reported in a monograph, *Ability Grouping and Scholastic Achievement*, by Svensson (1962). As pointed out above, the study took advantage of a unique situation. The City Council of Stockholm in 1955 decided to embark upon a program with a transitory organization of the 9-year compulsory school in the southern part of the city. Stockholm was thereby divided into two sectors. In brief, in the southern sector no separation whatsoever took place during the first 6 school years and in some districts not until the ninth grade. In the northern sector the dual or parallel system was still kept for a short time, which meant transfer for an "élite" to the selective academic school after the fourth grade. Both the northern and southern sections had similar social-class distributions. Since social background and initial ability measured by tests and school marks were kept under control, we could take advantage of the situation and carry out a 5-year follow-up. Comparisons were made between students of all ranges of ability in the two systems. By means of analysis of covariance we could compare outcomes taking initial social and achievement status into consideration.

Marklund (1962) carried out a study on class structure and class size. Since school marks all over Sweden were calibrated by means of achievement-test batteries in Swedish, mathematics, and English that are given uniformly to all students at certain grade levels, useful survey data were available for various research purposes. Marklund took advantage of these data. Homogeneity (as defined by IQ, school marks, and achievement tests) was studied keeping class size under control and vice versa, an approach that had not been taken by the numerous researchers who had tackled the problem of class size, for instance.

Marklund's study on the school class offers an interesting illustration of how previously evolved lines of communication to policymakers were used and reinforced. His monograph happened to be published a few weeks before the Education Subcommittee of the Budget Committee in Parliament had to consider a move made by some members of Parliament making provisions for a successive reduction of class size which, by the

way, at that time averaged about 25. The rationale spelled out in the motion was that such a reduction would increase educational productivity. Marklund was able to make it quite clear that within the range under consideration there was no evidence to support such a prediction, at least as far as student cognitive competence was concerned.

The role of the National Board of Education ought to be mentioned in this connection. It might be of interest to consider the budget requests that the National Board of Education in 1964 submitted to the government for the program of the Bureau of Research and Development which was instituted by Parliament in the 1962 Education Act. Parliament every year appropriates funds for its program. The Board then, on the basis of a policy framed as a result of discussions with its Advisory Council, allocates money to research institutions that have been approached by the Bureau or that have turned in proposals. In the budget request there is a lengthy discussion of the suitable percentage of the total educational operational budget to be allocated to educational research; 2% was by the mid-1960s suggested as a target for the near future (Svensson & Björklund, 1966). That target has not been achieved. By the mid-1970s it was still below .5%.

The Bureau in its budget requests for the fiscal year 1965-66 pointed out that educational research, in order to yield something that could be useful, should be centered around comprehensive programs carried out by research teams that can be kept together for several years. A research director has been assigned to each project, and the junior members of the team should be given the opportunity of utilizing their work as part of their graduate studies.

In conclusion, I see the future of research in education at three different levels and for three different purposes:

1. The universities should pursue their genuine and unique function, which is basic research. This does not imply a withdrawal into an ivory tower, but simply that their researchers should be given sufficient time and resources to develop the new concepts and models without which further research, particularly short-range and expensive developmental work, cannot be productive. It is not the role of the university scholar to develop instructional systems or introduce new classroom practices. His role is properly fulfilled if he stays aloof and indulges in the "irresponsible" thinking that precedes developmental work.

2. Large-scale surveys evaluating new organizational setups introduced by central or local authorities, or projects producing instructional systems which are expected to guide classroom practice, ought to be carried on outside the universities. Since quite a lot of this research deals with the evaluation of new structures or curricula, it is obviously important that the evaluation should be carried out by independent research bodies, such as

the National Foundation for Educational Research in England and Wales, or at the Max-Planck-Institut für Bildungsforschung in Berlin. Apart from the fact that, ethically, a governmental or quasi-governmental agency ought not to evaluate its own operations, research within a bureaucratic setting can easily become a nightmare to the researcher. His work is by definition antibureaucratic, informal, innovative, and not limited by timetables.

3. Obviously, certain administrative functions within the central or regional educational agencies cannot be successfully performed unless special research techniques are used to secure data. I am thinking primarily of the planning function which has to be guided by the collection of information on pupils, teachers, schools, and finance. In some instances, regular sample surveys have to be conducted. In recent years, research in some countries has focused on massive attempts to improve curricula and instructional methods. The tremendous increase of funds which the federal government of the United States is investing in improving school practices is well known to all of us. In countries where a central ministry of education more decisively influences the curriculum—I am not thinking particularly of England—one has begun to realize the possibilities of changing practices by preparing teacher aids or instructional systems. The sheer size of such an enterprise and its financial implications make it a task for a ministry.

We can draw up a continuum which is located between the two poles: on the one side, complete aloofness in the ivory tower, and, on the other, total immersion in the noise of actions and discussion of issues in the educational marketplace. The problem is that of striking a proper balance—of establishing a good liaison with the policymakers without becoming evangelists and without scaring the other partner into withdrawing his support.

2. ASSESSING RESERVES OF ABILITY AND PROVIDING EQUALITY OF OPPORTUNITY FOR ADVANCED EDUCATION

In the late 1940s the provision for equality of opportunity for advanced education (beyond mandatory schooling) became a major issue in Swedish educational policy. The parallel structure of the system with selective transfer to academic secondary education of the British grammar school or French *lycée* type after 4 (or sometimes 6) years of primary schooling was regarded as the major cause of the loss of talented students from lower social strata and thereby as a barrier against equality of opportunity.

This issue inspired research to determine how significant the reserves of ability were and in which social strata they were located. These studies

offer an interesting illustration of interplay between educational research and educational policymaking. It also illustrates how policy-oriented research comes up with methodological by-products of value for a particular branch of research that goes beyond its usefulness in a particular situation.

The Commission on Student Welfare, which was appointed shortly after the war, had to consider various measures to facilitate access to higher education for students from underprivileged homes. The Commission, chaired by Dr. Edenman, later Minister of Education and closely involved in the framing of school policy in the Labor Party, sponsored several studies. One of them was a statistical survey of how students reaching the matriculation examination and going on to the university were recruited from various social strata. This problem had previously been considered by the 1940 School Commission, which made a thorough analysis of the various types of academic secondary schools, with special reference to the 4- and the 5-year *realskola* (Statens Offentliga Utredningar, 1944).

The 1940 School Commission arrived at the percentages presented in Table 1, basing its studies on the enrollments in 1938 for the whole country.

TABLE 1
Social-class Background of Secondary-school Students in Sweden in 1938

Social class	Five-year <i>realskola</i> (%)	Four-year <i>realskola</i> (%)	Four-year municipal <i>realskola</i> (%)	Gymnasium (%)	Social class distribution among electorate (%)
Upper	18	9	7	28	5
Middle	58	55	50	61	38
Lower	24	36	43	11	57

The Student Welfare Commission sponsored a study of the social background of matriculants and university students carried out by S. Moberg and C. E. Quensel (1949) of the University of Lund. They arrived at the percentages of students registered for the matriculation examination and of those already matriculated at the university (Statens Offentliga Utredningar, 1948a, p. 31) as shown in Table 2. Moberg (1951), later Minister of Higher Education, carried out a thorough analysis of three age groups of matriculants and studied their university and occupational careers in relation to their social background and school marks.

In a follow-up study by the present author (Husén, 1969), when the entire male population of the city of Malmö was followed from the third

TABLE 2
Social-class Background of Secondary-school Matriculants
and University Students in Sweden by Mid-1940s
(Source: Moberg & Quensel, 1949)

Social class	Students registered for matriculation examination (%)	Students matriculated at the university (%)
Upper	51	58
Middle	42	36
Lower	7	6

grade of elementary school (age 10 years) to call-up for military service (age 20; see Husén, 1950), it could be shown that enrollment in the academic type of secondary school was very unequally distributed among the social strata. The percentages of further schooling within the four strata into which the population was divided are shown in Table 3.

Neymark (1961) has made an investigation based upon a 10% sample of the total age group of conscripts comprising practically all males born in 1928 in Sweden. The first series of data was collected in 1948 in connection with their call-up for military service. He determined the distribution, by father's occupation, of male students who had passed the matriculation examination. As can be seen from Table 4, the discrepancies are striking.

Findings like those presented in Table 4, showing a high overrepresentation of students from upper- or middle-class homes in the academic type of secondary school and among entrants to the university, gave rise to the question of the extent to which a talent reserve existed in the lower social strata. There was no reason to believe that the working class, which made up more than half the electorate, could not provide more than 10%

TABLE 3
Social-class Background of Students Entering Academic Secondary School
in the City of Malmö by 1948 (Source: Husén, 1950)

Social class	Further education in academic secondary school (%)	<i>Realskola</i> only (%)	<i>Gymnasium</i> (%)
Upper	85	24	61
Middle	40	30	10
Upper-lower	22	13	9
Lower-lower	12	9	3

TABLE 4
Distribution, by Father's Occupation, of Male Students
in Academic Secondary Schools in Sweden
(Source: Nymark, 1961)

Father's occupation	Total number of male students	Proportion having passed matriculation examination (%)
A. Farmer:		
(1) Farm owner	803	2
(2) Tenant farmer	141	2
B. Small-farm owner:		
(1) Farming, gardening, forestry, etc.	194	0
(2) Farm laborer with small holdings	169	1
C. Fisherman	53	0
D. Small merchant	120	11
E. Small manufacturer or artisan	85	9
F. Building contractor	76	9
G. Repairman	75	0
H. Taxi owner	45	2
I. Civil servant:		
(1) With university training	54	80
(2) Other specialized training	102	53
J. Top-level manager	36	70
K. Middle- or lower-level manager	44	23
L. Supervisor or technician:		
(1) In public service with assignment not requiring advanced general education	52	15
(2) Privately employed	69	8
(3) Foreman	88	3.5
M. Salesman	50	2
N. Worker in public services:		
(1) Fireman, policeman, coastguard, etc.	44	5
(2) Traffic worker (railroad, bus, transit), mailman	83	5
O. Skilled worker in industry:		
(1) Construction worker	199	2.5
(2) Craftsman (skilled carpenter, mechanic, etc.)	173	3
P. Semi- or unskilled construction worker:		
(1) In public services	68	1.5
(2) Road worker, railroad worker	79	0
(3) Construction worker (building industry)	133	0
Q. Industrial worker other than in categories O and P	488	1
R. Forestry worker	99	0
S. Farm laborer (other than B(2))	187	0.5
T. Worker in distribution and transport	120	0
U. General laborer	37	3

Note: Based on 10% sample of 20-year-olds born in 1928.

of the university students, while the upper class, with less than 10% of the adult population, was able to provide three to five times as many university students (depending upon the way of classifying the occupational groups into social strata).

Moberg (1951), who has devoted thorough studies to the problem of the social origin of and the type of professional training completed by matriculants in Sweden, has shown that the proportion of students from the lower social class passing the matriculation examination was fairly constant between 1910 and 1943. Even if we assume that social mobility to some extent might have brought about a higher level of potential (hereditary) scholastic aptitude in the middle and upper classes, the difference in enrollment between these two classes, on the one hand, and the lower, on the other, must be accounted for by other factors, such as the earning power of the family, the level of educational aspiration of the parents, and the attitudes towards higher education. Before proceeding to present some of the studies concerned with social factors preventing the utilization of talent, we shall take up the problem of how to estimate the size of the talent reserve or the potential academic population. The phrase *talent reserve* (or *reserve of ability*—*begåvningsreserv*) was coined in 1948 by the later Minister of Education, Dr. R. Edenman, who was the chairman of the Student Welfare Commission, when giving a talk to the general conference of university students at Uppsala. At that time he could substantiate his statements about the size of the talent reserve with findings from studies made by the present author (Husén, 1946, 1947, 1948).

In order to carry out an investigation aiming at an estimate of the size of the potential secondary-school or university population, two conditions have to be fulfilled. First, comparable measures of ability, such as intelligence test scores, achievement test scores, or standardized school marks, have to be available for representative groups. Second, population registers should allow follow-up studies for a number of years. The first condition is self-evident, since no dependable estimates can be made unless ability data for representative samples of individuals are available. An estimate must, however, also be built upon a follow-up of the sample in question from the point in the school career when differentiation begins until the point when it is finished. The whole age group of conscripts for military service, all 20-year-olds, provided the data for a comprehensive investigation in 1945 (Husén, 1946). Data were collected for an age group of 44,011 individuals, comprising more than 95% of the total male population born in 1925. The conscripts were given a group intelligence test composed of eight subtests, most of them verbal in content. The tested group was divided according to their formal education into the following groups: elementary school only; middle school (*realskola*), without leaving examination; middle school (*realskola*), with leaving

examination; university preparatory school (*gymnasium*), without leaving examination; university preparatory school (*gymnasium*), with leaving examination; and university studies.

The test score distributions for each of the subgroups were determined, as was a series of measures of spread, such as lowest decile, lowest quartile, average, highest quartile, and highest decile. The potential academically talented school population was then estimated in the following way. A comparison was made between those who had reached their middle school examination and those who had dropped out, mainly because of failure to meet the academic requirements. Only 50% of the intake in the *realskola* reached the leaving examination in the prescribed time. Another 10% to 15% repeated one or more grades. The proportion of pupils not being promoted at the end of the spring semester was, on the average, 10%. Assuming a "cutting point" at the lowest decile in the distribution for the group with leaving examinations from the *realskola*, we would then leave aside some of those with low intelligence who by means of coaching and other artificial means had survived the *realskola* and reached the leaving examination. A cutoff score identical with the lowest decile in the distribution of those who reached the leaving examination would divide the distribution of those without leaving examinations into two equal parts, which corresponded to the opinion held by some investigators that about half the failures in the *realskola* are caused by lack of school intelligence and the other half by nonintellectual factors. Now, applying the cutoff score we had derived from the distribution of the total age group and the subgroup with elementary school only, we found (Husén, 1946) that 33% of the total age group surpassed an IQ of 107, which then would imply that the *realskola* with its present requirements would fit the intellectually best endowed third of an age group. About 20% with only elementary school had scores above an IQ of 107.

Applying the same procedure to those who had entered the *gymnasium*, we found that 14% of the age group would have the intellectual assets to pass the matriculation examination (*studentexamen*).

We could then find out how many of the tested group with formal schooling below the cutoff score for the middle school examination and the *studentexamen*, respectively, would have the potential to profit from more advanced education. Thus, it turned out that less than half of those having these scores in the total age group actually had proceeded to more advanced education. Only about a third had reached the *studentexamen*, which accordingly indicated that there was a considerable potential university population.

The same methodology was applied to successive age groups from 1946 to 1948. The group intelligence test was completely changed from 1947 to 1948 but did not affect the percentages. The findings from the four age groups are presented in Table 5. Outcomes of the surveys of

TABLE 5
Estimated Number of Individuals
Capable of Passing Major Secondary-school Examinations,
Expressed as Percentage of Age Group
(Husén, 1946, 1947, 1948, 1975)

Age group	Individuals above test-score limit			
	For lower secondary school (<i>realskola</i>) examination		For university entrance examination (<i>studentexamen</i>)	
	Within total age group (%)	Among those with elementary school only (%)	Within total age group (%)	Among those with elementary school only (%)
1945	33	22	14	5
1946	32	20	13	5
1947	32	19	12	4
1948	33	22	13	5

entire age groups were quoted by the Minister of Education in the Education Bill on Comprehensive Education submitted to the Parliament in 1950 (Prop. No. 70).

Since, at the end of the 1940s, the actual number of students taking the matriculation examination was about 5% of the corresponding age group, the talent reserve would consist of at least as many as those actually passing the examination and admitted to the universities. The same conclusion applies to those who passed the lower secondary school examination. This would then mean that a tremendous talent reserve, mainly in the lower social classes, was not utilized for the type of occupations for which the academic schools give the basic preparation.

The methodology briefly described above for estimating the relative number of academically talented, and for assessing to what extent there exists a reserve of nonutilized talent, has several weaknesses. They will be taken up in turn.

It has been pointed out that the lack of reliability of the test instrument used as a measure of ability represents a source of error. The distribution of scores for those having completed a certain type of schooling and those who have not even entered upon it will overlap more than would the distributions of the "true scores," since the error variance increases the overlap. This would then mean that the number of those in the groups of lower schooling passing the critical limit will be greater with low test reliability.

A much more important source of error, however, has been pointed out and thoroughly analyzed by Ekman (1949). A group intelligence test

is far from perfectly correlated with either the level of formal schooling reached or the criteria of success in the academic secondary school. Under favorable circumstances, such as when the restriction of range effect is not too strong, the correlation ceiling is about $+0.70$, which means that about half of the variance in school performances can be accounted for by factors measured by intelligence tests. Lack of validity is an important source of error and will increase the degree of overlapping between the distributions of the subgroups in the test and therefore contribute to an overestimation of the size of the talent reserve. The overlapping effect will be much greater than that caused by lack of reliability.

The test was administered at the age of 20, that is, after the differentiation into various types of schooling had taken place. Those who had gone so far as to reach the *gymnasium*, or had passed the university admission examination, were being tested immediately after having finished school, whereas those who had left school after completing elementary school (ages 13 to 15) were at a disadvantage with regard to the kind of intellectual capacity measured by the test. In a 10-year follow-up study, it was shown that the academic type of schooling leading to the *studentexamen* improved the IQ by about 10 points (Husén, 1950). Hårnqvist (1959), employing a more adequate technique, estimated the rise in IQ to be about 15 points. This source of error would then tend to reduce the overlap between the distributions compared and would contribute to an underestimation of the intellectual reserve. It is therefore highly desirable to have means of indicating the level of ability before any differentiation in formal schooling has taken place.

On the basis of the criticism advanced by Ekman (1949, 1951), one can, following his suggestions, make these distinctions:

1. The "reserve" in the wide sense of the word should be distinguished from the "reserve" in a narrow sense. In talking about reserve in the broad sense, we include all the factors (cognitive and noncognitive) that are related to success and failure in the academic type of education. We have been able, in the studies mentioned above, to assess only some of the cognitive factors, and therefore our estimate is at best limited to the reserve in the narrow sense.

2. The "potential reserve" should be distinguished from the "actual reserve." When testing an age group at the age of 20, part of the individual differences is due to variations in formal schooling and different occupational experiences, the interval between the end of formal schooling and the time when the test was administered.

The ideal conditions for a study of the size of the talent reserve would then be to have data pertaining to intelligence and school performances at the age of 11 to 12 and then to follow this group for a period of years, in Sweden to the age of 20 to 21, when most students pass their university admission examination and when differentiation in formal schooling has been completed.

Härnqvist (1958) made an investigation on the reserve for higher education for the 1955 University Commission. In doing so, he was able to build upon theoretical conceptions advanced in the previous research and to utilize experiences gained in the administration of studies of this kind. But he has also made several new contributions to the methodology in this field, and his study is beyond any doubt the most valuable to date. Thanks to the efficiency of registration he could take out samples for which data were available at several points of time and did not have to wait for follow-up data.

The school records of two groups were studied, one completing the fourth grade of the elementary school in 1945 and the other completing in 1949 the second to the last grade (the eighth school year) in the *realskola*. The sampling technique used may be described as stratified cluster sampling with systematic selection within strata. The elementary-school group contained one-fourth of the classes in the fourth grade, a total of about 10,000 boys. (Since Härnqvist wanted to use the test data obtained at call-up to military service, this sample was confined to boys. The *realskola* group contained 3,600 boys and 2,800 girls.) School marks at the end of the fourth grade for the elementary school were available, and the same applied to the *realskola* group.

Härnqvist adopted a modification of a procedure proposed by Ekman (1951), which he called the probability method and which can be described as follows:

1. The assumption is made that all the children in the upper social class who possess the necessary minimum aptitude to profit from academic secondary schooling actually get the opportunity. Previous studies by Husén (1950) and Neymark (1952) had shown that about 85% of the upper-class children had had some secondary academic-type schooling, and about 60% to 65% of the total group had come as far as the university admission examination at 20. There were even some with an IQ below 100 who had succeeded in getting into the *gymnasium*. The probability that a student of a certain initial level will be transferred to the academic type of school is determined by the number of males at each ability level within the upper social class (social class 1) who actually transferred.

2. The number of pupils at various initial levels (according to school marks) is multiplied by the probability computed at each initial level from social class 1. This yields an estimate of the number of pupils in social classes 2 and 3 (middle and lower) who, given the opportunities of those in the upper class, would have the potential to be transferred to academic education.

3. The probability that a student, having embarked upon academic training, will pursue his education to a leaving examination is derived from data available for social class 1. The assumption is that in social

class 1 optimal conditions govern which keep the student in school and prevent him from dropping out.

4. The number of pupils in the middle and lower social classes, according to step 2 above, are multiplied by the probabilities arrived at according to step 3. This yields estimates of how many students there are in social classes 2 and 3 who have the abilities to pursue their studies to the leaving examination, given the opportunities of social class 1.

5. The numbers arrived at, according to step 4, are then summed up and reduced by the numbers of students who actually have reached the school-leaving examination in question. The rest will then be the estimate of the talent reserve. According to Ekman's terminology, we are then dealing with the "potential reserve in the wider sense."

Using the procedure briefly described above, Härnqvist arrived at the following results:

	Percent
1. Having actually passed university admission examination	8.2
2. Having passed leaving examination from <i>realskola</i> , but, according to estimates, capable of passing university admission examination	4.1
3. Having not even been transferred to <i>realskola</i> , but, according to estimates, capable of passing university admission examination	15.5
4. Having passed the leaving examination from <i>realskola</i> , but, according to estimates, not capable of passing university admission examination	6.5
5. Having not been transferred to <i>realskola</i> , but, according to estimates, capable of passing the leaving examination from <i>realskola</i>	17.1
6. Not capable of passing the leaving examination from <i>realskola</i>	48.6

According to the procedure applied by Härnqvist, about one-fourth of the boys by about 1955 would have the potential to pass the university admission examination within an age group in Sweden (categories 1 to 3 above). Since only about 8% actually passed the university admission examination, the potential university reserve in a wider sense would then be about one-fifth of the age group. Categories 1 to 5 inclusive would be capable of passing the leaving examination from the *realskola*. The potential reserve in the wider sense would be about one-third of the age group. On the whole, then, Sweden at the beginning of the 1950s did not develop even half of its intellectual potential, and in the case of the potential university population less than one-third.

The crucial problem in discussing the validity of the results obtained in the study referred to here is the extent to which the assumption is valid. One might question the assumption that, given the environment of the upper class, the young people in the middle and lower classes (with the same initial capacity in terms of school marks) would proceed to the academic type of secondary school and on the whole pursue their study there to the same extent as is actually the tendency among upper-class students. The assumption implies, in order to be valid, an almost egalitarian society that has existed for at least one generation. Research findings hitherto available have shown that the attitude towards higher education and the willingness to put the children into institutions of higher learning, or to have a relatively high level of aspiration, are not so strongly correlated with economic factors as with the educational background of the parents. As an example, by the end of the 1950s, 68% of the sons of university-trained people passed the admission examination to the university. The corresponding figure for the rest of the occupations referred to in social class I (upper) was 42%. (It should be mentioned that in the categorization scheme worked out by the Bureau of the Census, all professionals with university training belong to social class I.) This difference in frequency of university admission examinations in favor of the academically trained group should be related to the fact that the non-university-trained group in the upper class has a higher average income.

The release in early 1957 of the Härnqvist survey gave rise to an intensive debate among both researchers and policymakers. The former, of course, raised methodological issues, some of them related to the one briefly dealt with above. The latter, depending upon political faiths, either found the findings supportive or rejected them because they refuted their own convictions. The biggest newspaper, *Dagens Nyheter*, ran several editorials and on its opinions-editorials page the present author entered into polemics with the head editor who had authored one of the editorials. The findings had repercussions in the Ministry of Education, not the least because the Minister of Education was Chairman of the 1955 University Commission that had initiated the survey. It was used as supportive evidence in the main report of the Commission a few years later.

3. DIFFERENTIATION, INDIVIDUAL DIFFERENCES, AND DISPARITIES IN EDUCATIONAL OPPORTUNITIES

3.1 *Structure of Compulsory Schooling*

The term "differentiation" was pivotal in the Swedish debate on the reform of compulsory education from the late 1940s until the 9-year comprehensive school was definitively confirmed by an Act of Parliament in 1962. It concerned a central issue of the educational policy, namely, the integration of the elementary school and the academic secondary school

(Husén, 1962). The debate, however, was conducted mainly on the pedagogical plane. It was purportedly concerned with what was most "efficient": restricted ability range as in the selective academic secondary school or unrestricted ability range as in the elementary school.

The starting point of the debate on differentiation was the commonplace observation that students differ greatly in respect to abilities and interests. These differences are striking already when students enter school as infants, and they increase with time. When instruction in reading is begun, the 7-year-old Swedish students are found to range over three grades with regard to reading ability (Malmquist, 1962). Some children at the beginning of the first grade can read as well as the average students in the third grade. Others know the letters and can read, though haltingly; while a third category is quite ignorant of all linguistic symbols. At the end of the first year at school, some students are equal to average fourth-grade students, while others are still struggling to combine letters to make words. Variability increases in the higher grades. In the sixth grade, for example, reading skill has a range corresponding, generally speaking, to six grades. The same picture is found in other fields of knowledge and skills. If teaching is to be carried out successfully and if the demand is that each student be given an opportunity to develop according to his own potential, steps must be taken which take these differences into account. Such measures aim at promoting the students' individual development paying due attention to their different starting points. But in view of the aim mentioned here, instruction can hardly be intended to make the differences still greater; in any case, this should not be a primary goal. Measures connected with education and upbringing, and taking the varying abilities of pupils into account, have in Sweden been referred to as *differentiation* or individualization.

Differentiation measures may be organizational or pedagogical. The former have bearing upon either the division of students into different types of schools or grouping pupils into parallel classes within a certain track in accordance with some criterion—in most cases marks, teachers' ratings, or examination results. The latter refer to such measures within the framework of a given class which facilitate instruction and will give each child the greatest possible benefit, such as assigning students to subgroups within the classroom or using multilevel reading material.

Organizational differentiation includes the following measures:

1. *School type differentiation.* Students are distributed among different schools, often by selection to the most prestigious. Examples are the transfer which previously took place from the fourth grade of the elementary school to the first year of the 5-year lower secondary school (*realskola*) in Sweden, and in England the transfer from the primary school to the secondary grammar school (11+ examination).

2. *Track differentiation.* The students are assigned to different lines of

study in the same type of school. These tracks are kept more or less separate in respect to curricula and instruction. Examples of this kind of differentiation are the organization of Grade 9 in the comprehensive school or the various lines of the continuation schools.

3. *Class differentiation.* The students have, on the whole, the same curriculum, but are divided into parallel classes on a certain grade level according to some criterion of general ability, such as achievement, test scores, grade point average, or intelligence test scores. An example is streaming in British comprehensive schools (see Sjöstrand, 1961; Jackson, 1965). A somewhat "contaminated" type of class differentiation was the parallel arrangement in the seventh and eighth grades until 1962 in Sweden in most of the experimental comprehensive schools on the basis of the students' choices of subjects and alternative courses. This kind of agreement is in American and British educational literature often referred to as "homogeneous grouping" or "ability grouping" (Yates, 1966).

4. *Ability grouping* or "*setting*." The students within a given grade are grouped and regrouped in parallel classes in certain subjects, according to their progress in the respective subjects (see Conant, 1959). This means that they are regrouped for each subject. An example of ability grouping was the so-called Österåker system (named after a school district outside Stockholm where it was first introduced) in which students were regrouped, for instance, in English and mathematics according to ability, but studied other subjects together with their classmates. This kind of arrangement is sometimes referred to as "sectioning."

The measures referred to, at least in Sweden, as *pedagogical differentiation* should rather be called individualization. They are all concerned with something done within the framework of the class. Group instruction may be given, as in reading in the lower section (Grades 1-3) and English in the middle section (Grades 4-6). A class may be divided into halves according to ability in the subject in question, and the teacher instructs one half at a time. Such a division into groups may also be made by keeping the groups into which the class is divided in the classroom, or in the classroom and available group rooms, and allowing them to work in parallel. It is possible, particularly if self-instructional material is available, to give individual tasks on which each student can, in principle, work at his own speed and at his own level of difficulty.

3.2 *The Problem of Differentiation in*

Swedish Educational Reform, 1940-1960

In 1937, school attendance by Parliamentary Act was made mandatory for 7 years (instead of 6 years) in the whole country. Some local education authorities had already established 8-year elementary schools. This meant increased parallelism between the elementary school and the academic secondary school (*realskola*). This was one reason for the appointment of

the 1940 School Commission, which meant the initiation of the new reform era leading up to the 1962 Education Act regarding basic education and the 1964 Act regarding *gymnasium* education (Paulston, 1968).

The question of linking between primary and secondary school proved to be the most difficult problem this Commission had to tackle. The Commission was divided in its recommendations into two almost equal factions, one of which was in favor of the existing "double linking," while the other wanted a 6-year common primary school from which transfer could take place. A special report, "Linking Between the Elementary School and Higher Schools" (Statens Offentliga Utredningar, 1944), gave both the arguments and the facts on which both sides based their views. Among other things, the Commission followed up during a period of 8 years cohorts of altogether 10,000 students entering different types of *realskola*. Various criteria were used to evaluate the 4-year and 5-year *realskola*. Thus the frequencies of nonpromotion and dropout were determined, as well as the mean marks in the leaving examination and the average age at which the examination was taken. There were no great differences between the two types of *realskola*, except in social recruitment. Recruitment to the *realskola* connected with the *gymnasia* was more one-sided, and the greatest social representativity was found in the municipal middle schools.

The main arguments advanced by those wanting an *early* transfer to the *realskola*, i.e., after the fourth grade, were as follows:

1. Students would take the leaving examination or its equivalent earlier (about 6 months were saved).
2. Psychologists were of the opinion that academic aptitude can be diagnosed with sufficient accuracy when a child is 11 years of age.
3. Greater homogeneity from and including the fifth grade would be obtained in both elementary and secondary schools, to the advantage of both instruction and students.

The chief arguments in favor of a *late* transfer, i.e., after the sixth grade in the elementary school, were:

1. Aptitude, either academic or practical, cannot be determined with any great degree of accuracy in 11-year-old children. A differentiation of children of this age might easily be determined by irrelevant economic and social factors.
2. A common 6-year primary school would lay the foundation of understanding among the different social strata and contribute to the development of a sound civic spirit. "The system allowing special forms in *realskola* and girl schools parallel with the fifth and sixth grades of the elementary school may easily exaggerate social class tensions in our society" (Statens Offentliga Utredningar, 1944, p. 225).

The School Commission asked for the views of the four professors of education in Sweden and published their reports in a supplement entitled

"The Present Standpoint of Psychological Research in Respect of the Mental Growth of Children and Young People, etc." (Statens Offentliga Utredningar, 1943). The professors were unanimous in the view that it is possible to select the more gifted 11-year-olds for academic studies. One of the experts on the Commission, Ejnar Neymark, who submitted a special report with the title "The Guidance of Young People to Education and Vocation" (Statens Offentliga Utredningar, 1945), claimed that the psychological conditions necessary for individual vocational guidance did not exist in girls below the age of 14 to 15 years or in boys below 15 to 16, i.e., at the earliest in the eighth grade. If it were only a question of selection for grammar school studies, Neymark claimed that it did not matter whether transfer to the *realskola* took place after the fourth or the sixth grade. At both levels, the young people were too immature to have any realistic idea of their own potentialities.

Some members of the Commission drew the conclusion from the expert report that since suitability for academic studies was dependent mainly upon general intelligence, and this could be determined satisfactorily at the age of about 11, transfer should take place at that age. Others, however, came to the conclusion that if the academically gifted students went to higher schools early, the elementary school would be deprived of its most gifted students. To be on the safe side, students who were both practically and theoretically gifted would probably move over to the grammar schools. This argument became of great importance for the 1946 School Commission.

This School Commission was appointed in 1946 and submitted its main recommendation in 1948 (Statens Offentliga Utredningar, 1948b). It recommended a Solomonic solution: The parallel types of schools were simply to be incorporated into one unitary, compulsory school. The first 9 years at school would be common to all children and form a "comprehensive school." No organizational differentiation was to take place before Grade 9. This grade was to be divided into three tracks, one to prepare students for the *gymnasium* (9g), one general (9a), and one a preparatory vocational (9y) track.

The School Commission, after much research into, among other things, the development and structure of ability, came to the same conclusion as the 1940 School Commission, namely, that academic ability can be diagnosed as early as a child's 12th year (11+; Elmgren, 1952). Two conclusions were drawn from these studies: (a) that practical and theoretical ability are positively correlated, and (b) that practical ability is more complex than theoretical ability and that its components tend to mature later than those which make up theoretical ability. If students were selected for an academic education at that age, not only the clear-cut academically gifted students would be included, but also the best of those in whom good theoretical gifts and good practical gifts are combined.

"Since it may be assumed that the great majority of the students selected, or at least the best of them, would be led by way of academic tracks to the *gymnasia* and higher education, an early differentiation of this kind would eventually deprive the practical vocations of academically gifted people. The practical occupations, and the social classes engaged in such occupations, would suffer a depreciation, a split would occur in society and democracy would be jeopardized" (Statens Offentliga Utredningar, 1948b, 69ff.). The outcome of the Commission's research, it was said, favored a late differentiation and a flexible system of alternative choices.

The Commission's recommendations were submitted to a large number of public agencies and private organizations for criticism and review, a practice which is typical of the commission system in Sweden. Before the government decides to draft legislation, commission reports are submitted to a wide selection of evaluating bodies. Among these were 160 local boards of education, 44 out of 51 state school inspectors, and more than 200 faculties at *gymnasia* and lower secondary schools. All the faculties but three were opposed to the suggested organizational solution, which was symptomatic of the attitude of the secondary-school teachers to the proposed organization. The risk of reducing the efficiency of instruction by having ungrouped heterogeneous classes in the upper grades was pointed out consistently. It was feared that this would lead to a lower standard of education.

On the basis of the Commission's recommendations and the observations made by the evaluating bodies, a bill was drafted and submitted to Parliament in 1950, entitled "Principles Governing the Development of the Swedish Educational System" (Prop. No. 70). One cardinal point in the bill was the problem of "uniformity and differentiation." The Minister of Education began by quoting a declaration made by the 1940 School Commission that "an organically integrated school system comprising all forms of schools, in which each growing individual, regardless of residence place and of the socioeconomic status of the parents . . . will be given, if necessary with public assistance, an education suited to his aptitude."

The Minister added:

Work of a reform intended to bridge old social gaps must ensure that the school system for all classes of society appears as a uniform whole, in which lines of study are available to all young Swedish people, and where every growing individual, regardless of his social starting-point in life, can learn how to make the best of his talents. Such an aim is incompatible with any kind of parallel school system. Differentiation into separate schools should, I believe, not be made until it is necessary on account of choice of vocation. (p. 190)

It was pointed out in the bill that the School Commission's recommendation to individualize instruction in undifferentiated classes had been adversely criticized by teachers and others. This was perhaps only natural in view of the lack of suitable methods and teaching aids.

The discussion of the problem of differentiation has now been carried so far that I do not think investigation will carry us much further. The problem must now be solved by practical experiments. The ten years ahead of us before the educational reform can be realized to any great extent should be used to determine the methods of instruction available, to test the capacity of the different designs, and to arrive at final conclusions regarding the organization of the comprehensive schools on the debatable points. (191ff.)

The decision made by Parliament in 1950 anticipated a rather long period of pilot work with the 9-year comprehensive school, under the supervision of a special division of the National Board of Education. Permission (including grants) was granted by the government to 14 local boards of education to commence pilot work. Local education authorities were generally eager to start experimenting with the new school, and applications for permission had to be sifted strictly.

Since there were *realskola* in the same or nearby communities, there was from Grade 5 on a considerable transfer of students from the comprehensive school to other types of schools. This transfer amounted, according to calculations made by the National Board of Education, to about 20% of the students in the last three grades of compulsory schooling (13 to 16 years of age). This meant that the upper section of the experimental schools had to work without most academically oriented students (cf. National Board of Education, p. 200; Statens Offentliga Utredningar, 1961, p. 78). Furthermore, the supply of teachers was inadequate. It was not until 1955 that the experimental schools were authorized to engage subject teachers with tenure in the upper sections, and then there was in Sweden a growing shortage particularly of subject teachers.

The National Board of Education, in its evaluation of the School Commission's recommendations (Statens Offentliga Utredningar, 1949, p. 212), emphasized that the results achieved by the experimental schools were to be studied carefully, and "compared continuously" with those of the *realskola*. The so-called comparison philosophy, which was endorsed by the Parliament and used as a formula to achieve consensus about the 1950 bill, led to a large part of the experimental division's time being devoted to comparing the knowledge imparted by the comprehensive school and *realskola*. These comparisons played a very important role in the progress reports of the experimental work each year submitted to the government by the National Board of Education. Apart from the fact that such evaluations can only incompletely measure how far the aims of the school—other than those of pure acquisition of knowledge—had been realized, the public did not take into account that the "new" school had certain aims quite different from those of the "old" schools. In the sometimes heated debates on the effectiveness of the experimental schools, it was forgotten that a school's work cannot be judged by the final products alone (cf. Husén, 1960, p. 70 et seq.). In the *realskola*, only about 50% of

the students reached their goal, the leaving certificate, without delay, while nonpromotion was comparatively rare in the comprehensive and elementary school. Also the price paid for the intellectual quality of the final product had to be taken into account. From the beginning the great majority of the teachers working in the upper section were in favor of an organizational differentiation which would make possible the grouping of students according to options in Grades 7 and 8.

The great problem during the experimental work of the 1950s may be said to have been the creation of satisfactory flexibility in the choice of subjects and tracks, in spite of the demand made by many educators for a division of students into theoretically and practically oriented.

The question of differentiation proved to be the cardinal issue for the 1957 School Commission, as it had been for the 1940 School Commission and the 1946 Commission. The 1957 Commission, which had to prepare legislation for an introduction in the whole country of the comprehensive school on the basis of the experiences gained from the pilot program, based its opinion on, among other things, a rather extensive research program and on the report of the 10 years of experimental work published by the National Board of Education. Brief reports of some of the investigations are given below.

The Commission pointed out in its main report that differentiation may be referred to in terms of two systems, a selective and an elective system. By selective system is meant in principle that society or the school, successively or more definitively at a certain point of time, makes a decision regarding admission to a certain type of education. Thus the decision is not the concern of the student or his parents. The elective system, whether the choice is made definitively at a fixed time or successively, implies that the student, together with his parents, may, within a given framework, choose the education deemed suitable, regardless of the school's opinion of the student's qualifications. This system is based on the principle that the school provides information to enable parents and students to make their choices (Statens Offentliga Utredningar, 1961, p. 254).

In its main report, the Commission gives certain rationales for its recommendations.

1. The students should pass through the upper section on the basis of their own free, successive choices: "The principle of free choice is one of the fundamental principles of the new school" (p. 255). A classification of students by the school would be a violation of the fundamental democratic principle of self-realization.

2. The comprehensive school must give "a more extensive and a more varied education than the previous school gave" (p. 258). The common base of general education must be increased. Then the common frame of

reference will be widened for all members of society, and a better foundation will be laid for specialization.

3. The school must promote the students' social development and create better understanding between the different social classes. The school must be designed to encourage training in consideration for others, community spirit, and cooperation. "It is desired to create a democratic school environment giving all students equal opportunities to develop their personalities and to become good citizens; great care must be taken before students are separated from each other at a certain age and sent either to different schools or to different tracks of study in one and the same type of school" (p. 260).

4. The need for education in rural areas should be satisfied at a level equal to that of urban areas. It would be difficult to combine organizational differentiation with this principle, since it would imply that many children from rural areas would have to go to some other place to get the education they wanted.

5. The disparity of prestige of different occupations should be diminished. This can be done by practical vocational guidance and by instruction in civics.

6. The efficiency of the school should be increased as far as the encouragement of ability, interest, and ambition in young people is concerned. This does not mean, as one might be tempted to believe from the discussion of differentiation, taking care of the intellectually inclined students only, but fostering all types and degrees of ability. A selection at an early age tends to be more highly correlated with social background than a later selection. Society will take better care of the reserves of talent by not making any classifications of students at an early stage.

The Commission proposed a more diversified system of courses in the upper section of the comprehensive school than was available in the 1950s. These consisted partly of a common nucleus of compulsory subjects, partly of a choice of groups of subjects. No fewer than nine tracks were proposed for Grade 9 instead of the former three.

3.3 *Differentiation and Social Background*

The respective merits and drawbacks of comprehensive and selective schools have often been discussed with reference only to the *end products* they turn out. In other words, comparisons have been restricted to the quality of senior-class students in both types of schools. Most of the comparisons are limited to intellectual achievement; a few, like Miller (1961), include other educational outcomes. Two attributes of the traditional European academic school have enabled it to supply the universities with young people of outstanding intellectual talent: (a) its students are largely *selected* on the basis of scholastic ability to begin with, and (b) it has

been in a position to thoroughly screen out weaker students during the course of their schooling.

Screening has until recently been an important element in European secondary education and applies both to transfers from one level to another and to the pursuit of studies within a given level. In West Germany, selection for the *Gymnasium* until recently took place at 10, but in most provinces (*Laender*) less than one-fourth of those admitted take the final examination (*Abitur*), in spite of severe competitive selection from the beginning (Undeutsch, 1955). In France, between 40% and 50% failed in the corresponding examination (*Baccalaureat*), despite all the screening that has occurred up to and during the *lycée* stage (Husén, 1960). In Britain, the committee appointed by the Ministry of Education to look into grammar-school failures noted that *more than one-third* of the students never got as far as a final certificate, in spite of their having been screened on admission by means of the 11+ examinations (*Early Leaving*, 1954). In Sweden, the incidence of grade-repeating and failures was studied by the 1957 School Commission (Orring, 1959).

In the light of these empirical data, we might well ask how this screening process is supposed to function. This question has relevance both for selection from one level to another and particularly for grade-repeating and dropouts. It stands to reason that a selected and screened-out student body is more homogeneous in regard to scholastic ability and as such is easier for the teacher to deal with. This in turn makes it possible to attain a higher standard of achievement. However, the "efficiency" of a school system must be assessed with the compass of a "strategy," that is, a total evaluation in which *all* student categories are included. This means that we cannot consider only that part of the educational raw material which emerges in fully processed shape in accordance with a certain standard model. We must also take into account how much of the raw material remains either unprocessed or never gets beyond the semimanufactured stage.

The issue may be formulated thus: *What price has to be paid for the high standard achieved by a few?*

The selective, academic school is imagined to achieve its fine results with its end products by all the time restricting itself to a selection limited solely to ability or motivation. Those who are not selected or who are screened out are considered to be either less gifted or less industrious than their classmates. It is taken as axiomatic that everyone is competing on equal terms, that is, everyone is offered equal opportunity. Ability and energy alone are the determinants of scholastic success, not social background. If social handicaps are present, they are to be removed by financial aids of one kind or another.

Ability and school motivation admittedly carry great weight, but they are far from being the sole determinants. In competitive selection, wide latitude is allowed to social factors, perhaps not first and foremost the

economic status of parents, but rather their level of education and social aspirations; this turns out to be surprisingly true, even in societies like the Swedish where economically based class differences are fairly small.

A large body of research (see Husén, 1975) has shown that students from socioeconomic group 3 (consisting for the most part of working-class families) or from certain categories of socioeconomic group 2 (for instance, agricultural families) are greatly underrepresented in the academic secondary school.¹ In France, for example, until recently only one-eighth of the student population came from working-class homes, which form two-thirds of the nation's families. Similar findings are reported by Dahrendorf (1964) for West Germany. When Premier Khrushchev in 1958 announced the new educational reform in a speech before the Supreme Soviet, he said that 60% of the enrollment in Moscow's upper level schools was recruited from the intelligentsia and bureaucracy and only 40% from the homes of manual workers and farmers (Conant, 1959).

A study of the social-class origins of Swedish undergraduates by Moberg (1951) found that roughly the same class proportions had existed from 1910 to 1943.

Boalt (1947) investigated enrollment by social class in 1936 in academic secondary schools in Stockholm. A similar study pertaining to Stockholm, Gothenburg, and Malmö in 1938 was made by the 1940 School Commission (Statens Offentliga Utredningar, 1944). In 1955, the last year that the dual system was operating in Stockholm, the School of Education surveyed the social background of the students of Stockholm's North Side who in that year transferred to a *realskola* from Grade 4.

The same picture emerges from both studies: Over a long period of economic leveling out, the conditions of recruitment have altered only slightly. It might well be asked: Do the observed disparities between social classes derive from the selection mechanisms as such, so that students from a lower social class are handicapped in competitive selection from one level to another? Boalt (1947) worked out a technique for estimating the degree of social handicap which asserts itself in selection from a lower to a higher school type and in screening within a school type. He defined social handicap as the relationship observable between selection (or screening) and social factors, with scholastic ability held constant.

As part of its attempts to evaluate the learning outcomes of the 4-year and 5-year *realskola*, the 1940 School Commission (Statens Offentliga Utredningar, 1944) made an extensive study of dropouts and grade-repeating in both school types. As already mentioned, the Commission followed the progress of an entire age group for a period of 8 years,

¹ The socioeconomic grouping employed here is the one developed for the Swedish Election Statistics. The three groups could roughly be labeled "upper," "middle," and "lower" class, respectively. "Social class," "social group," and "socioeconomic group" are used interchangeably.

beginning with the first grade in the *realskola*. It was found that only about half the students passed the final examination (*real.xamen*) and entered the first year of the *gymnasium* without repeating any grades; in other words, they reached the study objective within the period envisaged by the instructional program. When students who had repeated grades for one or more years were included, it was found that less than two-thirds reached the study objective (Statens Offentliga Utredningar, 1944, p. 325). A similar study was carried out by Orring (1959) for the 1957 School Commission. He sampled one-fifth of the admission quotas for a given year and was provided with data on the progress of students on up through the grades. The study objective was attained within the regular period by about the same proportion of students as 17 years earlier, when the admitted portion of an age group had been much lower. More than one-fifth of the students in the 5-year *realskola* did not graduate or transfer to the *gymnasium*; the corresponding proportion for the 4-year *realskola* was almost one-third. Orring also related scholastic progress to marks received in the fourth and sixth grades of elementary school, since these had been employed as criteria of admission to the *realskola*. He found the incidence of dropouts and grade-repeating to correlate significantly with the aggregate marks at time of admission; however, dropouts were also considerable among students having high aggregate marks.

Orring's data have been supplemented by Husén (1966) with reference to the social-class background of students. He studied the relationship between dropout and social class. It was found that 60% of the students from socioeconomic group 1 reached the final grade in the 5-year *realskola* without delays, as compared with about 48% of the students from socioeconomic group 3. Similar social-class differences are discernible as well for the 4-year *realskola*. We thus observed that the mechanism of social selection, which Boalt was able to establish for one age group of students in Stockholm, applied to Sweden as a whole.

Research into the effects of organizational differentiation, conducted by Svensson (1962), made it possible to study the relation between social class and competitive selection for academic secondary schooling. This particular study bore reference to 5,960 students of Stockholm's North Side, who in the spring of 1955 competed for entry into a higher school. The schools supplied information on marks and occupation of parents. Students were given a group intelligence test as well as the standard battery of achievement tests at the end of the fourth grade. Three different measures of ability—marks, achievement test scores, and intelligence test scores—were related to outcomes of the competitive selection considered in terms of social background. With three measures of ability under control one could show how accepted applicants and rejected applicants were distributed by social background.

To begin with, it will be noted that those admitted to the academic

secondary school show quite a different social-class distribution from those who continued in the fifth grade of elementary school.

Since a much larger proportion of students is admitted from socioeconomic group 1 than from group 3, it might be expected that group 1 also has the larger proportion of rejected applicants. But that is not the case. The proportion of rejects in each of the three social classes was:

Socioeconomic groups	1	2	3
Rejects in percent	31.9	44.2	42.7

In the following we shall venture to explain the mechanisms that might underlie this handicap for applicants from group 3.

Theory says that a competitive selection shall be made solely on the basis of ability. Among the criteria set up for this purpose are "intelligence tests," most of which are constructed to measure scholastic ability. Their validity is established by correlating test scores with school marks. In the Stockholm study, scores achieved in the group intelligence test were drawn upon to divide the students into nine ability groups. The top 4% were put into group 9, the next best 7% into group 8, and so on. Group 5 included students who clustered around the average, while other groups fell below the average (cf. Svensson, 1962).

There were fewer applicants from socioeconomic group 3 than from groups 1 and 2 at all ability levels. However, this difference varies with ability level in that students of average or below-average ability from group 1 apply about three times as often for admission, whereas among the 11% who scored highest on the test, the number of applications from students in group 1 slightly exceeds that of the other classes. We thus see that the very intention of applying for admission correlates significantly not only with the student's ability but also with his social background.

With these facts in mind, we might expect a higher number of rejected applicants from groups 1 and 2 than from group 3, since it could be reasonably assumed that the larger number of applicants from the first two classes (and group 1 in particular) would contain more students who were less able and less qualified for school studies. But the number of rejects tended to be larger in group 3, in spite of the much lower incidence of applications. This tendency was less pronounced among the most able students than for those who clustered around the average. We may accordingly say that social handicap is operative not only in building up the level of aspiration, but also when this aspiration is translated into action, that is, when students from the cultural environments represented in group 3 apply for an academic education.

At this point several objections could be raised: First, an intelligence test does not measure factors relevant to scholastic ability, or at best measures these only incompletely; second, tests of this kind favor certain

groups of students. The latter objection is based on the observation that an ability test tends to discriminate against students from a lower social class. This suggests that the social handicap we noted above in selection to advanced education would have been even more pronounced if we had been able to use a more culture-free intelligence test. The second of these two objections may therefore be said to reinforce rather than weaken our conclusions on social handicap. In answer to the first objection, the Stockholm study did not exclusively rely on the intelligence test but could also draw on the results of achievement tests given in mathematics and Swedish. On the basis of scores achieved in eight tests, students were classified into nine ability levels according to the same procedure as above. Finally, we had access to school marks, which of course comprise an instrument whose use is held to permit the making of a fair selection.

On the whole, the same picture emerged of the selection mechanisms operating in the transfer from primary to academic secondary school when achievement tests and grades were employed as criteria instead of IQ. Students from a lower social class at a given level of ability according to all three criteria consistently showed lower frequencies of application and higher proportions of rejections.

The study of social background and transfer to academic secondary school (Husén, 1960) was a by-product of the larger survey on the comparative survey conducted in Stockholm from 1955 through 1959 (Svensson, 1962). The purpose that the 1957 School Commission (and initially the researchers themselves) had in mind was to elucidate the effects of a selective and a comprehensive school structure on student achievements.

In the fall of 1959, when I was at the University of Chicago as a visiting professor, letters from my institute conveyed the message that the School Commission, which was supposed to recommend a more permanent structure of the mandatory school on the basis of the experiences gained during the extensive 10-year pilot stage of the 1950s, had reached a stalemate. Evaluations of the pilot program by a committee of which I was a member (*Försöksverksamheten*; National Board of Education, 1959), as well as ongoing research, did not provide unequivocal support for either of the two systems, the selective or the comprehensive, in terms of student competence at the end of the lower secondary stage. The social-democratic members of the Commission wanted an undifferentiated basic 9-year program, whereas the more conservative members wanted differentiation according to academic criteria after 6 years of schooling.

I submitted two memoranda to the Commission's secretariat, one presenting the findings of the study of enrollment and social background referred to above and one on criteria of evaluation. During the entire 1950s, under the spell of the "comparison philosophy," it had been taken for granted that the ultimate criterion of what system was "superior" was

the competence of the students that came out of the respective systems. Social selectivity and loss of talent in such a process was seldom, if ever, made explicit. One now began to realize that equality of educational opportunity was not always compatible with efficiency. Furthermore, one began to realize that efficiency had to be viewed in the context of the overall costs. What price is paid in terms of inequality and loss of talent by employing selection mechanisms of which dropout and grade-repeating were at least as powerful as the selection from one stage to the next by marks and examinations.

Thus, the researchers contributed to the creation of an awareness among policymakers about the social implications of various solutions to school structures. As can be seen from the Parliamentary record in 1962, when the Comprehensive Education Act was debated on the floor, the reform was considered more a social than a pedagogical one.²

4. OBJECTIVES AND CONTENT OF CURRICULA

4.1 *Studies for the 1957 School Commission*

The Swedish Government in 1957 appointed a commission of inquiry, over which the Minister of Education presided, to make recommendations for reforms which were later incorporated in the 1962 Education Act on the 9-year comprehensive school (*Grundskolan*). Among the directives of this body was to "conduct required investigations into the objectives of the school and its purposes in society." Further, "Curriculum contents and methods have . . . from time to time to be reassessed as to their purposefulness towards the introduction of the young generation into society." Concurrent with the appointment of the 1957 School Commission, the Industrial Council for Social and Economic Studies (SNS), an organization sponsored by Swedish industry and business, launched a research program designed to thoroughly elucidate the curriculum content of core subject areas in the last three grades of the 9-year comprehensive school.²

This was the first time that empirical methods were employed in Sweden to explore the relevance of certain aspects of formal education for the requirements of the job world, as well as for the subjectively felt needs of individuals in their occupations and outside of them. In scope

² In order to give non-Swedish readers, not least those in the United States, a proper perspective on the Swedish educational scene, the following conditions should be considered. We are here dealing with a country with a fairly homogeneous population of 8 million and with a rather centralized and uniform school administration. The curricula, as far as they imply expenditures in terms of teacher salaries and class sizes, are decided upon by the Parliament. The content of instruction with regard to major topics and curricular elements is approved by the Government. Textbook materials are approved by a central agency. Teacher salaries and capital provisions in constructing school facilities are uniform. In a small country the relevant groups—in this case policymakers and educational researchers—are relatively small and opportunities for contacts frequently occur outside the formal settings. This makes it easier to establish communications between parties involved in educational change.

and methodology it might be said that some 20 years ago the program was unique even on the international scene. The Institute of Educational Research at the School of Education in Stockholm was called upon to take the main responsibility for the design and conduct of this analysis of subject matter in certain areas of school learning. While SNS desired to put particular emphasis upon implications for business and industrial occupations, the study was not limited to "marketable skills" but sought also to find out what the individual "needs" of citizens are in a technological democratic society.

In the first phase of the study, primary attention was given to the "three R's." Data were sought that would contribute to the revision of the curriculum in the upper stage of the new 9-year school. Grades 7 through 9 were regarded as critical. There had been little previous experience with a secondary school enrolling all children, for students of this age had been mainly those attending select academic secondary schools.

When the 1957 School Commission began its work, the SNS project proved to be well adapted to inclusion, and joint sponsorship was arranged. At the same time the project was extended to include physics, chemistry, and civics. The first report, on mathematics and Swedish, was published in 1960 (Dahllöf, 1960; Husén & Dahllöf, 1960). A report on physics and chemistry came out in 1962 (Johanson, 1962). Civics, finally, was reported on by Bromsjö in 1965.

4.2 Aims and Limitations of the Studies

A curriculum, it has been said, should answer four questions: Why? what? when? and how? The research for the 1957 School Commission, being mainly concerned with curriculum content, was by design limited to give a partial answer to the question, What? The intention was to obtain as complete a picture as possible of the requirements put upon individuals within and outside the work situation in terms of specific skills and pieces of knowledge. Our approach to curriculum study was "positivistic," though with full awareness of the shortcomings of this approach that are often mentioned: narrow utilitarianism, slighting of humanistic values, atomistic or mechanistic conceptions of learning.

It would in principle appear to be a simple task to find an answer to the question: What mathematics and communication skills are required in a certain occupation? Some would think it easy to answer this question without reference to actual school curricula or methods of teaching. It was soon learned, however, that the study would be sterile without considering what was actually happening in the classroom. It was found necessary to express the requirements for individuals in the work situation in terms of skills and knowledge imparted to the students or aimed at by teachers, that is, in terms of elements of the curriculum. In order to be of relevance to those who build curricula, the findings had to be communicated to

these educational craftsmen in their own language.

The demands that various types of vocational schools impose upon the students coming to them from schools providing general education had to be expressed in terms of curriculum topics in order to have relevance for teaching and had to be set not only against what the school does in striving to produce certain skills or knowledge but also against the outcomes of teaching as measured by expert judges and objective tests.

Three objectives for the investigation were set down at an early stage:

1. To form a concrete picture of *actual teaching practices* within the given subject areas and school types, that is, to shed light on instruction procedures and methods, the use of teaching aids, and the allocation of classroom time to the different elements or topics of a subject.
2. To survey in detail the *knowledge and skills* required of the individual both as holder of an occupation and as a citizen, based on the formalized ratings of teachers, work supervisors, and the employees themselves.
3. To determine *learning outcomes*, both immediate and long-term, by means of (a) the systematic ratings of teachers in receiving schools and of supervisors and employees, and (b) objective achievement tests given both at the end of school and a number of years afterwards.

In short, one wanted to examine both what individuals are taught and what they learn in relation to adult behavior and performance. In practice one had to relate three types of data to each other: first, the time spent and methods used in teaching the elements of a subject-matter area; second, the rated or tested outcomes of this teaching; and, third, the adult needs for certain skills and pieces of knowledge.

The significance of this investigation will be clearer if its limitations in methods, populations studied, and aspects of curriculum studied are set forth in some detail.

1. The study, in its first phase, was restricted to the three R's. These skills are normally regarded as the most important cognitive outcome of basic school instruction. Panels of curriculum or subject-matter experts assisted in drawing up the topic inventories. Only those topics could be included which were accessible to available measurement techniques. Hence, oral communication was omitted. The particular skills had to be stated concretely and illustrations provided. There was no pretense of including all outcomes of instruction within the subject-matter areas included for study. Noncognitive outcomes in the form of attitudes, interests, and values were omitted, which of course was an important limitation.

2. Only students in Grades 7 through 9 of the comprehensive school or parallel school types were included. But the measured outcomes of instruction evidently include the total effect of the 9- or 10-year course. This segment received primary attention because of the pressing task of designing a secondary-school program for all children. Ample experience

exists for the first six grades, which in many places have been taught within a unified pattern; but in Sweden there was a rather limited experience with a compulsory, i.e., universal, secondary-school program.

3. The adults studied were limited to those aged 18 to 30, for one main reason. If broader age groups had been studied, they would have been rather heterogeneous in terms of the amount or kind of schooling they had received. With this limitation on age range, adequate samples could be obtained from the sectors of employment we wanted to study.

4. The adult samples from various occupational areas included only personnel having no formal schooling above the lower-secondary academic level (*realskola*). To include the total hierarchy of employees would have necessitated including a study of *gymnasium* and university education. Since we were interested mainly in obtaining guides to a revision of the compulsory secondary-school curriculum, we dealt only with employees in the manual and lower clerical categories.

5. While we were sensitive to the impact of technological and other social changes upon individual lives, we chose to take an essentially static point of view. There are reasons to assume that these changes, even over decades, will have relatively limited effect on the basic skills (the three R's); the learning effects in these areas are more generalized and applicable to a wide range of life situations.

6. Finally, it was decided to procure a representative sample for the nation with respect to school teachers and employees in business and industry. Teachers were chosen by stratified random sampling. For employees this method (even cluster sampling) would have yielded very small samples spread over the whole area of Sweden, a procedure which would have been prohibitive, administratively and in cost. Instead, areas typical of Swedish employment were chosen and then enterprises sufficiently large to provide an opportunity for group testing were chosen by utilizing centrally available statistical data on size and occupational structure of enterprises. For details readers are referred to Dahllöf (1960).

4.3 *Assessing Course Content in Grades 7 Through 10*

The first step was a careful analysis of curriculum syllabi and textbooks in order to arrive at a manageable number of content elements or topics, such as decimal fractions, triangles, nonfiction reading comprehension, and dictionaries. These elements are given in general terms in the curriculum manuals issued by the National Board of Education, the topics to be covered at a certain grade level being enumerated in a few lines. The textbooks are structured according to these elements. The task of the expert panels we consulted was to break down course content into elements that could be regarded as relatively distinct. In this task use was made also of committees from educational agencies and teacher associa-

tions at both elementary and secondary levels. The opportunity of examining subject matter and methods for a given content area as a structured whole proved to be a fruitful experience for these people. Many teachers spontaneously reported gaining a new perspective on their work when they had to fill out the questionnaires derived from these expert dissections of course content. The teachers were stimulated in looking back on how they had taught a certain class the previous year and had to make a fresh evaluation of their expenditure of time and energy on different parts of the subject matter.

In order to make use of the experiences of a national sample of teachers, an elaborate questionnaire was developed that would require them to set down time spent, methods used, outcomes, etc., for various teaching units. In order to facilitate filling out the questionnaire in mathematics, an inventory of mathematical operations and problems was constructed using representative items from textbooks.

An elaborate procedure was worked out to motivate teachers to collaborate, for the authorities did not regard this work as part of the teachers' official duties. Nevertheless, the National Board did send out a circular to school principals indicating that special compensation might be offered, such as relief from certain classes. Since the project was sponsored by the government, the Board gave active support in other respects and the Director General of the Board signed letters sent out emphasizing the importance of the study. Letters were sent also by the chairmen of the national teachers' associations soliciting full cooperation. The research point of view was set forth in a letter by the project director to explain the purposes of the study and to guarantee anonymity. Some 2,000 mathematics teachers and 1,800 Swedish language teachers were included in the first phase of the study. Eighty percent or more responded.

4.4 Assessing Admission Requirements of Secondary Schools

After completing their formal general schooling, at about 16 years of age, approximately half of the students went directly into the labor market. The others entered vocational schools or general, college-preparatory *gymnasias*. The first group was approached directly in the field. The children continuing their schooling further were not contacted directly; their teachers were asked to evaluate the qualifications of students in the subject-matter areas especially considered.

All of the *gymnasium* classes in Swedish and in mathematics were included; the response frequencies from these teachers were 86 and 88, respectively. Letters from the chairmen of the appropriate teachers' associations were sent to each teacher.

There were about 500 vocational schools, some public, some proprietary (providing vocational training in some 900 areas), and all were included in the sample; again, special letters were sent to each teacher.

The percentage of teachers replying was about 85 for most types of vocational schools, a little higher or lower in some cases.

Both *gymnasium* and vocational-school teachers evaluated the quality of preparation their students had received in their previous school. The report was to be based upon the specific class taught in order to limit tendencies toward generalized answers. The replies were mostly in the form of ratings on a 5-point scale. The mathematics curriculum, for example, was divided into 48 topics. Each respondent indicated to what extent a certain topic was considered necessary as a preparation for the particular vocational course and to what extent students did or did not possess the desired facility.

4.5 Assessing the Requirements of Employment and the Individual Needs for Specific Skills

The field study of employed persons had two aims. We wanted to ascertain in detail the requirements for subprofessional personnel (both blue and white collar) by employers and to match employees' actual skills against these requirements. The first aim was satisfied, by no means perfectly, by having supervisors rate specific needs and also by asking workers to rate how often they made use of certain skills as well as indicate skills they needed but which they did not possess with their present training received in school. The employees had to indicate by means of the same technique to what extent they made use of the specific skills during their leisure time, that is, as citizens in a democratic society. The most time-consuming part of the field project was the measurement by standardized achievement tests of what workers had retained from school.

For each of the principal branches of the national economy, occupational distributions were obtained from the files of the Employers Association. Those occupational groups were included for which employees were located in clusters large enough for group testing and for which job classifications were unambiguous. The individual enterprises within each branch were ranked according to numbers of employees in the categories desired and the largest ones were taken. Where there was a numerous group of some other included occupation employed in a particular firm, they also were included. In the wood industry a group of small enterprises clustered by area had to be used. The total sample tested in mathematics and Swedish included 48 firms and 1,586 employees.

Since the firms in effect donated one day's pay for each employee included, and since some employees were expected to object to the invasion of their privacy, participation was solicited. The national associations of employers and of labor unions gave their support to the study and wrote to individual firms and unions in support of the project. In addition, local executives and union leaders were briefed on the study prior to

testing in each plant. The industry distribution of firms studied was as follows: mining, shipbuilding, metal, wood, food processing, and finance. Of all employees approached, 91% participated, each spending 5 to 6 hours. Of some 230 supervisors asked to make ratings (similar to those given by teachers in vocational schools), 96% cooperated.

4.6 Establishment of Reference Norms

Most of the data obtained in these various parts of the investigation hitherto reported were either observational or formalized ratings. The only way to test the reliability of the ratings was to compute interteacher correlations of ratings. There was, of course, no way of determining whether the ratings represented the "real" experiences of the respondents. Moreover, one might suspect that the ratings made by supervisors or by teachers in the vocational schools or in *gymnasia* merely reflected the opinions of the raters rather than the factual requirements. But opinions about standards are themselves important facts that determine the fate of students being admitted to school or while in the school, as they also determine the hiring policy and training programs of enterprises.

It was desirable to procure objective measurements with which to check the ratings and establish reference norms for the study of retained knowledge. This comparison was made possible by the achievement tests given to employees. A more fruitful interpretation could be made of the test results if one could establish some kind of measure derived from the school situation. It would be desirable to use the same tests for both workers and for graduating students. Our main interest was in two reference groups: those having left school after 7 years of compulsory school, *folkshola*, and those having left after passing the last grade of the first stage of academic secondary school, *realskola*. (Where possible, we selected firms in localities where basic schooling until recently had been only 7 years.) The same tests as were given to workers were taken also by representative samples of 980 students in the seventh grade of *folkshola* and by 618 students in the graduating class of *realskola*.

A meaningful interpretation of the test results achieved by the various occupational groups could, as has been pointed out, be made only if reference norms could be established. Therefore, the same tests were given to graduating students in the types of schools in which most of the employees tested had received their basic education. But the crucial problem in this connection was the comparability of the student group and the adult group. Since we intended to assess the retention of skills in the three R's, we were obliged to compare the adults with graduating students of the same level of proficiency as they themselves once had shown. We had no data for the adults at school-leaving age, but the school marks for the majority of them were available. The problem of comparability is then reduced to the question of whether the marking practice had remained

unchanged during the interval between leaving school and the testing of retained skills (between 5 and 15 years). We were able to check on this, since the school marks had during the last 15 years been standardized by achievement tests in the three R's which were administered to all students on a certain grade level each year. Therefore, the school marks could be used as control variables and comparisons made between students and adults by means of analysis of covariance. Thus, we determined the regression between marks and test scores in the student group and then compared the expected values according to marks in the adult group with their actual test performances. The difference between expected and observed values then provided us with a measure of the extent to which skills had increased or decreased after leaving school. The development of skills was in this way determined for various occupational groups and comparisons were made between employees with and without regular vocational training.

4.7 Results in Mathematics

The results from the two subject-matter areas, namely, mathematics and communication skills (Swedish), can be presented each according to the four methods adopted. First, the receiving schools, vocational as well as academic, had stated their *requirements as to previous knowledge* ("grounding") by means of a formalized rating technique. They had also judged the standard of knowledge of the student material they received. The ratings were made separately for each of the 48 and 47 topics or subskills, respectively, into which the experts had divided the course content of the subject-matter areas under consideration. Second, the teachers of the vocational schools were requested to estimate the extent to which each of the topics was a *prerequisite for the occupations* for which the vocational schools were preparing their students. The same types of estimations were performed by the supervisors and employees of the enterprises participating in the study. Since they had only a limited experience of the school curriculum and we had no opportunity to make the topics more tangible for them by means of concrete examples, we had to confine their ratings to a reduced number of broader topics. Therefore, the supervisors as well as the employees estimated the *frequency* of the use they made of the various skills. The supervisors, too, rated the proficiency the employees displayed in skills relevant to their jobs. The employees made, however, frequency ratings pertaining only to the same topics as those tapped by objective tests, thus a rather limited number of elements. Furthermore, they also rated the relevance of the various topics for their spare-time activities and for their functions as citizens in present-day society. In this connection they were requested to evaluate their actual skills, to determine the extent to which they felt they had shortcomings. Third, six subareas each in mathematics and Swedish were tested by means of

standardized achievement tests. The same tests were given to the employees and to representative samples of graduating students. Thus we were able to assess the development of some important skills after leaving school. Fourth, the actual instruction time devoted to each of the topics was mapped out by the teachers in the basic school by means of a formalized rating technique.

4.8 *The Need for Mathematical Knowledge for Further Vocational Training and Continued Academic Schooling*

The main findings related to requirements set up by the receiving schools can be summed up as follows, distinguishing four categories of topics:

1. Topics needed in all types of courses and all types of schools. These comprise the "minimum essentials" of mathematics; for example: arithmetic, common fractions, measurements, mental arithmetic, estimations, and applied mathematics (percentage, interest, and means).

2. Topics needed merely in specific courses; for example: applied geometry, simple business mathematics, and simple equations.

3. Topics needed in certain groups only in the *gymnasias*: geometry (theorems and constructions), difficult equations, algebra, square roots, the remaining part of applied geometry, and business mathematics of a more specific type.

4. Topics not needed; for example: slide rules, calculating machines (the study was conducted in 1957-1960).

Thus, it is possible to point out a "common denominator" or group of core topics that are basic to practically all types of training. The vocational training schools differed from the *gymnasias* in that the *gymnasias* put more stress upon components that are prerequisites for taking more mathematics, e.g., algebra and theoretical geometry.

The main results related to the teachers' ratings of the standard of the pupils' previous knowledge can be summed up as follows. In general, all teachers emphasized the lack of background in mental arithmetic, estimations, and arithmetic with common fractions. The vocational teachers, especially those of industrial schools, also emphasized the lack of preparation in many other topics—applied geometry, among others. The teachers of the *gymnasias* were not satisfied with theoretical geometry (theorems), equations, and algebra.

The "consumer" investigation in mathematics disclosed that certain topics, which may be regarded as fundamentals, are prerequisites for all kinds of further education, both vocational and general. These topics are: integers, decimal fractions and compound fractions, mental arithmetic and estimations, percentages, denominations, discount and interest, proportions, and averages. Certain topics are required only in specific instances. For example, applied geometry is confined to the industrial and

agricultural groups. Working problems, mixtures, and business arithmetic are likewise relevant for certain vocational curricula. Geometric theorems, square roots, algebra, and more complex linear equations are prerequisites for the *gymnasium* only.

There are thus, as far as previous knowledge is concerned, two distinct patterns of requirements, one advanced by the vocational schools and emphasizing the applications of basic operations and one set up by the *gymnasium* putting more stress upon components that are prerequisites for further study in mathematics.

4.9 *The Need for Mathematical Skills in Various Vocations and in Everyday Activities*

The need for mathematical skills in various occupations was rated both by teachers of the vocational schools and by the supervisors at the enterprises included in the field study. The correlations between the two sets of ratings tended to be fairly high, which is an indication of relevance. Two features stand out when making an overall consideration of the findings. First, the requirements as to mathematical skills for various occupations are on the whole very low and comprise a rather limited number of components of the mathematics course. Second, the complaints are fairly rare, which is a logical outcome of the fact that the requirements are very low.

The requirements for the vocational areas considered here are on the whole confined to operations with integers, common or decimal fractions, measurements, percentages, and measuring devices. There are practically no requirements in topics like operations with common fractions, theoretical geometry (theorems), and equations.

The requirements that the employees themselves had experienced during their spare time or in pursuing their activities as citizens were limited to three skills, namely, approximations and mathematical judgment, decimal fractions, and percentages.

4.10 *Retention of Skills at Graduation and Some Years Afterwards*

Standardized tests were given to 1,586 adults employed in various vocational areas and to students in the graduating classes of both the 7-year elementary school and the *realskola*. We wanted to assess the development of mathematical skills after leaving school in relation to formal basic schooling, vocational training, and occupation.

The school marks in mathematics were used as a control variable. Standardized achievement tests have been used in Sweden since 1944 as a means of creating a nationally uniform marking practice. The same written examinations have been given in all school units of the *realskola*. There were indications that the marks given in the middle of the 1940s are

approximately comparable with those given at the end of the 1950s. On the basis of the regression of marks on test scores in the graduating group, expected scores could be computed for the adult group. This value takes due consideration of the difference in distribution of marks between the school group and the adult group. Expected values have been compared with observed ones. Technically, by means of analysis of covariance, we have corrected the expected value with due consideration to differences in school marks.

Standardized achievement tests tapping six areas of skill were administered, namely, decimal fractions, common fractions, equations, percentages, geometry (items pertaining to the triangle), and estimations.

Employees with no formal schooling beyond the 7-year elementary school showed a clear-cut decline in skill in operations with decimal and common fractions, especially in the latter. In simple equations and applied geometry their results in Grade 7 of the elementary school are so very low that it is not possible to measure any loss of knowledge with the tests used. On the other hand, a considerable increase takes place in the two remaining elements, percentages and approximations including mental arithmetic. It should be kept in mind that these two elements have a high degree of usefulness both within the various vocations and for mastering quantitative problems that the ordinary citizen encounters in everyday situations.

Employees with elementary school plus some vocational training show increased skill throughout—with the exception of decimal and, especially, common fractions.

Employees with diplomas from *realskola* show on the whole a decrease of skill in all the elements except estimations. The decrease is especially noticeable in equations. It should be considered that the achievement tests had a certain "ceiling effect," which affected the difference between adults with and without diplomas from *realskola*.

Thus we found that skills with little use consistently display a more or less strong decrease in contrast to components that, according to receiving school supervisors and the employees themselves, exhibit a high degree of usefulness. Adults are far superior to graduating students in making approximations, estimations, and mathematical judgments.

4.11 *Implications for the Mathematics Curriculum*

The results obtained were compared with the course prescriptions for the various school types.

Certain conclusions for curriculum revision in mathematics were drawn and the following four problems were brought to the attention of the curriculum subcommittee of the 1957 School Commission. Its recommendations were then further elaborated by expert committees set up by

the National Board of Education and finally embodied in the Curriculum for the Comprehensive School issued for the entire country by the National Board of Education.

1. *Can we say that the mathematics courses of the various basic school types have a "common denominator" or a "common core" of skills or topics? If so, what are these topics?*

As already pointed out, there are about a dozen topics required for all types of training, both academic and purely vocational. The topics are, according to the rank order of their importance in requirements: operations with whole numbers, decimal fractions, measurements in the decadic system, approximations and mental arithmetic, percentages, *regula de tri*, interest, discount, and average. When speaking of a core course in mathematics we are then entitled to these elements.

2. *If, as was the case at the upper stage of the comprehensive school in Sweden, we have two alternative courses, of which the one is "theoretically geared" and prepares students for further academic training and the other is more "practically geared" and prepares students for vocational schools, is it possible to delineate the elements which belong to the one and to the other?*

The common core or denominator is, as has just been pointed out, considerable. Irrespective of the number of alternative courses considered, the above-mentioned elements are the common core making up the basic course in mathematics. If, as was the case with the comprehensive school, there is a more theoretically geared course preparing students for the *gymnasium*, the following elements have to be included in this course (besides the core elements): more complicated linear equations, square roots, algebra, theoretical geometry (certain theorems), planimetry, stereometry, and applied arithmetic.

In the more practically geared course, which prepares students for vocational schools, one has in the first place to consider the following elements (in addition to the common core): applied arithmetic, simple planimetry, stereometry (cylinder and prism only), uniform movement, work problems, mixtures, and very simple linear equations (introduced already during the previous stage of the elementary school).

3. *Are the elements of the "common core" (basic course) in the present school learned to a sufficient level of proficiency so that the quantitative tasks in adult life can be coped with?*

The achievement tests administered to graduating students and to adults indicate that some important topics are not mastered. The receiving school complains about certain basic weaknesses. Operations with whole numbers are apparently carried out with sufficient skill, but skills in other important elements are not satisfactory.

Irrespective of approach, we have found that increased attention must be paid to training in approximations, mathematical judgment, and mental arithmetic. These elements have a high degree of usefulness in the

everyday quantitative situations in most occupations. There is a demand for increased training for those bound for the *gymnasium* as well as for those bound for the vocational schools.

Fractions, especially common fractions, need to be learned more thoroughly. A surprisingly great number of students leave school apparently without sufficient grasp of the basic notions related to common fractions. Some of these do not understand the meaning of the symbols. It would appear that there is a need for a revision of the instructional content in this area. A good deal of the instruction has a doubtful relevance for practical situations.

Applied geometry must be more strongly emphasized. This holds true especially for the simple descriptive planimetry. Attempts to focus more upon the most important elements in applied geometry are especially recommended for the practically inclined students.

Further training in measurements and transformations of measurements is needed in order to prepare for the vocational schools in particular.

4. *What should in the first place be "weeded out" in case the instruction time has to be restructured to reduce certain topics or leave them out completely?*

Topics with low "usability" for the majority and that are requirements only for the *gymnasium* could be left out from the alternative course preparing students for the vocations. Such topics are, for example, algebra, which was taught in the practically geared alternative course in the comprehensive school, and theoretical geometry for the same category of students. Certain planimetric and stereometric elements might also be left out or reduced or moved to higher grades, for example, in the *gymnasia*.

4.12 *The Communication Skills as a Prerequisite for Vocational Training and Continued Academic Schooling*

The subject-matter area subsumed by the word "Swedish" was divided up into 47 topics or subskills which were distributed among eight branches, viz., reading, skills in handling books and other sources of information, oral communication, vocabulary, composition, grammar, spelling, punctuation, and Danish and Norwegian. The receiving schools rated their requirements as to the grounding in the elementary school on a 5-point scale. The standard for their students was rated according to the same technique.

In the strict sense the formal aspects of communication skills tend to be prominent in both the *gymnasia* and the vocational schools. In the rank order of mean requirements the following topics occupy the four most prominent positions in the *gymnasium*: use of the period, sentence structure, inflections, and spelling. The corresponding four elements in the vocational schools were, according to rank order: spelling, reading of

nonfiction prose, speed reading, and use of the period.

The elements of communication skills can be divided into three categories with regard to the requirements for further training set up by the higher schools:

1. Elements that to a high degree are required in *all* types of further schooling, academic and vocational: spelling, ability to grasp facts and main idea in nonfiction prose, rapid grasp of main idea in the paragraph by glancing through the material, and punctuation (proper use of the period).

2. Other elements that are required for *all*—but not to such a considerable extent as above—types of training:

Oral communication: clarity of speech, preciseness of pronunciation, narration (e.g., telling about an incident), demonstration (e.g., demonstrating an apparatus or a diagram)

Vocabulary: common words

Grammar: writing with correct syntax, producing fluent writing

Composition: filling in forms, narration (e.g., telling about an incident), description (e.g., describing an apparatus)

Punctuation: comma, period, etc.

Skills in handling books: use of dictionaries, encyclopedias, and similar reference books

Reading: understanding and critically evaluating points of view in nonfiction prose.

There is ample reason to regard the elements under these first two categories as the basic course in the mother tongue.

4.13 *The Need for Communication Skills in Selected Occupational Areas and in Everyday Activities*

It was, for various reasons, impossible to have the supervisors and employees rate the same 47 topics as did the teachers of Swedish. Some topics had clear connotations for teachers only. Therefore, in rating frequency of use, the elements had to some extent to be reformulated to make sense to those assessing communication skills pertaining to the working sphere. The supervisors thus had to rate 24 topics. Since the employees took a standardized achievement test covering four main categories of topics, their ratings of frequency of use were limited to these areas, viz., reading vocabulary, written expression, and oral expression.

All three categories of judges (vocational teachers, supervisors, and the employees themselves) familiar with the working conditions within the different vocations emphasized reading skills and oral expression much more than various types of written expression.

A structure analysis showed that the industrial vocations had rather

high intercorrelations with regard to the structure of requirements. Office and bank work differed markedly from the other types of vocations considered in this study.

The employees throughout emphasized the reading of short paragraphs as the communication skill most needed in their work. The supervisors indicated in their assessment of the standard of skills of their subordinates that certain aspects of oral expression were not satisfactory, for example, ability to give an account of a procedure.

The spare-time use of communication skills seems for large groups, especially for industrial workers, to be more important than their use in the on-duty situations. Certain sex differences within the same vocations pertaining to spare-time reading habits were noticed. Men more often read newspapers, journals, and books. Women read more magazines and wrote more letters.

4.14 *Retention of Skills at Graduation and Some Years Later*

As part of the field investigation covering some 50 enterprises, six standardized achievement tests were administered to the 1,586 employees 18 to 30 years of age. The same tests were also given to graduating students in the 7-year elementary school and the academic secondary *realskola*. Since the school marks were controlled, we could estimate the retention of the skills measured during some years after leaving school. The tests covered the following areas: reading comprehension, vocabulary, book skills, syntax, spelling, and punctuation (use of the period). They were modifications of tests taken from series of standardized achievement tests used as a means of bringing about a uniform marking system in Grade 6 in Swedish schools. The first three instruments measured, in a wide sense, reading skills; the last three measured writing skills.

Adults having only elementary school (7 years) as their basic formal schooling tend on the whole to achieve more than a comparable group at the end of the elementary school. Achievement on the vocabulary test is much higher among adults. Skills in handling alphabetic lists have improved, too. Reading comprehension shows no change. Spelling has improved considerably and punctuation to a slight extent, while syntax has remained unchanged.

No statistically significant differences were found between sexes, groups of different vocational training, or occupations. In this respect communication skills differ strongly from mathematics, where some elements display large differences between categories of training and vocations. The most plausible explanation for this fact is apparently that communication skills on the whole are much more generalized than mathematical skills, which are more connected with specific requirements.

4.15 *Implications for Curriculum Revision in Swedish*

By closer examination of curriculum manuals and teaching aids it is evident that much more emphasis is nowadays put on reading of nonfiction prose. The use of silent reading techniques had begun to replace the oral reading drill. Requirements stated by the receiving schools and representatives of the occupation studies clearly indicate that nonfiction prose occupies an important place in everyday vocational reading. This should be considered in connection with the fact that fiction prose and poetry occupied a dominant position especially in the *realskola*.

Leading specialists in the teaching of Swedish have in recent years emphasized the importance of learning the technique of how to study. Apart from its importance for real-life situations, it is of great importance for mastering the home assignments and the ordinary school work. We have found that around half of the teachers of Swedish are giving no formal instruction at all in techniques of study.

Skills in handling books and other sources of information have been emphasized more and more in the current course directions, for example, the use of dictionaries, encyclopedias, and library catalogues and the reading of maps and diagrams. Only about 1% to 2% of the total instruction time, however, is devoted to special training in book skills.

There is ample evidence that oral communications have a key role to play in most enterprises. Teachers at vocational schools and supervisors in industrial enterprises emphasize the importance of correct oral expression. Recent revisions of curricula in Sweden have given a much more prominent place to special training in oral expression than previous course directions did. The topics belonging to oral expression occupy about 10% of the instruction time.

Systematic improvement of vocabulary has gained a more prominent status in recent curricula in both the elementary and the academic secondary schools. But special practice in vocabulary improvement still occupies a background position, since it covers only about 1% in the elementary school and 3% to 4% in the academic secondary school. The very fact that the adult group scored considerably higher on the vocabulary test than a comparable group of graduating students is an indication of the usefulness of vocabulary skill. The receiving schools fairly consistently complain about the limited vocabulary of the students.

Compositions used to play an important role in the teaching of communication skills in Sweden. They are a kind of examination in that the students have to choose a topic from a list and write a composition within a couple of hours under supervision. The preparation and especially the going through of the corrected written work took considerable time, altogether about one-third of the total instruction time. The new curricula

according to the recommendations of the experts put emphasis upon short compositions written mainly as means of practicing written expression and not as examinations. As could be seen from our study of how the instruction time is used, short practice-type compositions were almost nonexistent in schools investigated here.

4.16 *Inquiries into Gymnasium (Pre-university) Education*

Dahllöf was in charge of two main studies for the 1960 *Gymnasium* Commission. The first one sought to clarify the achievement and proficiency standards which universities and colleges established for their first-year students and to pinpoint the sectors in which students allegedly fell short of these standards. The second study addressed the same questions to representatives of the working community in business and public administration. In both studies Dahllöf had to cope with two major problems: first, how to establish a core of general-education requirements in the form of "expectations catering to the public interest," i.e., a common "core of knowledge and skills required for all kinds of activity after completion of *gymnasium* studies, whether it pertains to further education, gainful employment, leisure pursuits, or the exercise of citizenship"; second, how to go about getting certain groups of "consumers" to specify the prerequisites they wanted.

Use was made of an extensive questionnaire in which the curriculum content of the *gymnasium* was broken down into 65 broad subdivisions. The respondents had to indicate the prerequisites applicable to their specific field and to rate the achievement and proficiency standards of *gymnasium* graduates. The technique employed with respect to the universities and colleges was to include the entire population, that is, responses were solicited from *all* the subject specialists, as a rule, the university full professors. These in 1962 numbered 570, of whom 528, or 93%, replied. It was found that the vast majority of the subjects taught at universities and colleges drew on a core of the *gymnasium's* curriculum content in virtually all university disciplines as defined by departments. The subdivisions that constitute such a core also come in for most of the complaints expressed in regard to achievement and proficiency standards. The respondents were asked not only to rate the prerequisites laid down for each *gymnasium* subdivision but also to give their judgment of its general educational value, independent of its value as a prerequisite in a specific subject-matter field. The findings are shown in Figure 1.

Each topic was rated twice on a 5-point scale. The first rating pertained to the importance of the topic as a preparation for studies in the discipline that the rater represented. The second rating concerned the general educational value of the topic (to what extent it pertained to "other activities in today's societal and cultural life which the student ought to be

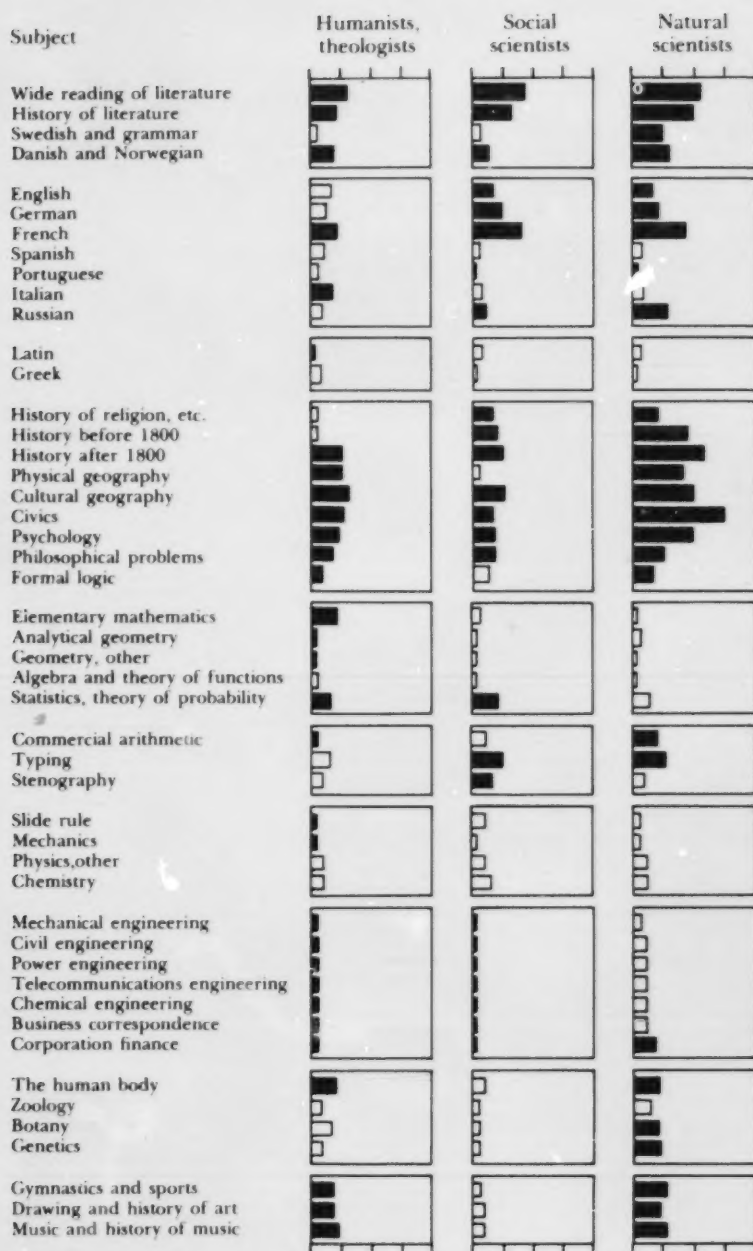


Figure 1. The contribution to general education of the various topics or subject branches of the curriculum in the Swedish *gymnasium* according to ratings performed by all the university professors in the country.

prepared to share"). The ratings were quantified by figures which had the following meaning:

- 4 = to a very high extent,
- 3 = to a high extent,
- 2 = in some important connections,
- 1 = in some not very important connections,
- 0 = not at all or to a slight extent.

The bars in Figure 1 indicate the *differences* between the medians of the two ratings, i.e., the difference between the value of a topic for general education and its value as a specific preparation for studies in certain disciplines. The differences that amount to more than .25 points on the 5-point scale are indicated by filled bars. These, then, give a measure of the "independent general education value" that can be attached to each topic.

We can see that the humanities and social science subdivisions are most highly rated for their general educational value. Scientists look upon the humanities as a complementary adjunct to their own education, while humanists do not take a corresponding view of the sciences. They give a rather low rating to the general educational value of orientation to science and technology.

Representatives of higher education were also asked to give views on the general aims and working methods of the *gymnasium* by means of both ratings and open-ended comments on these ratings. A large sector of opinion favored greater independence in the pursuit of *gymnasium* studies. "Broader general education and less specialization in curricula and departments, more self-activity on the part of students, and examinations covering greater segments of subject matter at one time rather than the hearing of lessons" (Dahllöf, 1963, p. 103). These suggestions agreed fairly well on the whole with those expressed in a concurrent study of matriculation candidates carried out by Härnqvist and Grahm, who asked the matriculants the same questions on working methods.

The cluster of subjects and skills in the *gymnasium* that form a common core of prerequisites for all paths of higher education turned out to be:

- (1) study skills, such as ability to read and take notes, and ability to do independent work, such as collecting, screening, and processing of information;
- (2) modern languages, principally English but also German, and, at a lower level, French;
- (3) elementary mathematics and statistics including theory of probability;
- (4) certain elements of the orientation subjects, especially in the social and behavioral sciences, such as political science, psychology, cultural

geography, and economic geography, and, to some extent, the history of thought.

The failings cited by university teachers pertained to skills in French and German (the latter to a lesser degree), study skills, statistics, social studies, psychology, and philosophy. These subject areas largely constitute the core of what may be regarded as the general education offered by the *gymnasium*. The complaints voiced by representatives of business and public affairs accorded on essential points with those of the academic community.

The Gymnasium Commission, which as usual among its membership had leading politicians, such as the subsequent Minister of Education and Prime Minister Olof Palme, observed these findings and by and large endorsed their implications in terms of curriculum development.

4.17 *Some Observations in Retrospect*

First, the research can help in tidying up problems as they are first encountered by the policymaker. Thirty-three years ago, shortly after I had completed my doctor's degree with a thesis on adolescent problems, I was invited to speak at a parent-teacher association meeting. After the lecture, when questions were invited from the floor, one of the audience asked: "Tell me, doctor, how late does developmental psychology think that it is appropriate for a fourteen-year-old girl to be out in the evening?" Not too different was a question that the Swedish 1940 School Committee submitted to the professors of education in Sweden: "From the point of view of developmental psychology, at what age would it be advisable to introduce teaching of the first foreign language?" As a young assistant my view did not count, but it was very simple: "Certainly by the age of two, the age when all children in England begin to learn English!"

The problems which are posed are often not only too broadly stated and, therefore, in need of being broken down but, in most cases, they also need thorough reformulation if they are to be backed by research methods. Thus, for instance, in studying the relation between age and language teaching, the cardinal problem is not age but methods of instruction. In 1941, the professors of education replied that the age of 11 would seem most appropriate, thus tacitly assuming that the traditional grammar approach should be used. Since then we have in Sweden started to teach all children English at the age of 10 and even at 9, a decision which some decades ago would have been regarded as a pedagogical sacrilege.

The comparisons between the comprehensive and the selective school structure in Sweden during most of the 1950s were naively carried out in terms of the conventional end products. Thus, the "grammar school" students in both structures were compared without consideration of the

fact that, at least partly, they pursued different objectives and that they differed considerably in terms of grade-repeaters and dropouts. They were compared, in fact, without reference to differences in the price paid for the end products. Therefore, it seems to me to be of great importance to establish at least occasional contact between researchers and policymakers. The dialogue between them should help to create more articulate conceptions of the problems and their implications.

Let me once again make reference to the unique situation in the city of Stockholm in 1955. The City Council decided to introduce the 9-year comprehensive system on a pilot basis in the southern half of the city. The dual or parallel system was kept in the northern part for a transitory period. This provided researchers with an opportunity to follow an age cohort over a 5-year period and to carry out continuous evaluations of the two systems, keeping social background and initial scholastic ability under control. This study was sponsored by the 1957 Royal Commission which dealt with the structure and curriculum of the comprehensive school. As the results from the evaluation emerged at the successive grade levels, they were reported to the Committee, with the Minister of Education in the chair and representatives of political parties among members.

Second, more closely at hand and certainly more obvious than the tidying-up task is the "know-how" or technical craftsmanship that researchers are expected to bring to bear on problems that the administrator or policymaker with (or, as a rule, without) the help of the researchers is trying to formulate. The researcher is called upon to do the fact-finding, to carry out the surveys that clarify the descriptive picture of a certain educational phenomenon. One fully appreciates the technique the researcher is applying when, for instance, he carries out sample surveys. The various Royal Commissions in England, such as the Robbins and the Plowden Committees, are cases in point.

One of my co-workers, Sixten Marklund (1962), carried out a study on the relationship between the size and homogeneity of the class and the outcome of instruction. National survey data were used, and for the first time, size and homogeneity were studied simultaneously: The effect of size was analyzed, keeping homogeneity under control, and vice versa. The monograph that reported the study happened to be submitted a few weeks before the Budget Committee of the Parliament was reviewing an amendment to the Education Bill made by some of its members and proposing a successive reduction of class size, which, by the way, in Sweden was and still is one of the lowest in the world. The argument for the motion was that a reduction of size would improve the outcome of instruction, a rationale which clearly was refuted by the study within the class-size interval considered in this particular case, i.e., between 20 and 30 pupils.

Third, the researcher can be called upon to assist in interpreting data which have already been collected. As part of the routine administrative procedures within all ministries of education, hosts of statistical information are collected. These data are particularly useful in improving planning within the ministry. The researcher can assist not only in interpreting data already gathered but also in advising what might be collected, so that more relevant aspects of the educational reality are covered by the statistics.

In recent years the planning function has grown rapidly to such an extent that research has become part and parcel of planning. I am not thinking primarily of fancy mathematical models which seem to shun encounters with empirical reality, but rather of the basic surveys that are needed.

In my own country as well as in other countries, I have encountered some confusion about the proper role of the researcher in the decision-making process. For instance, it has been argued that decisions about the school structure should not be made on the basis of "pure political deliberations." When in 1962 the Swedish Parliament passed legislation which made the introduction of the 9-year, undifferentiated comprehensive school mandatory all over the country, the objection was raised that this was a "political" and not an "expert" decision. A considerable number of the secondary-school teachers who had been trained for an elitist system resented the fact that their "expert" opinion was overruled. A colleague of mine in education, referring to that opposition, contended that the research evidence for the new setup was either not available or even contradictory. It seems to me that he and others had in mind some kind of Plato's Republic of kings, perhaps university professors, who ought to have the final and decisive word in public affairs.

It has often been stated that wars are too important to be run by generals. It has less often been emphasized that schools are too important to be run by educators, whether teachers, school administrators, or professors of education. The appropriate role of the educator in the decision-making process is primarily to provide information and expert judgments and thereby to help broaden the base upon which decisions are made by those who, in a representative democratic system, have been elected to make these decisions. If the educator is a researcher, it is even more important that he should play his role according to the rules and confine it to professional performance as long as he wears his researcher's hat. Let me hasten to add that the educator, whether a teacher or a researcher, is obviously entitled as a private person to advance opinions or make political statements. Equally self-evidently, however, he should try to keep these personal statements clearly separate from the generalizations drawn from his scholarly inquiries.

The nature of the research can easily affect the relationship between

the researcher and the policymaker. Problems pertaining to school structure are, since they have profound implications for the educational opportunities of groups of individuals, issues which easily can cause political temperatures to rise and governments to be shaken. This brings considerable pressure to bear upon the researcher. The directive given to him might be confined to the study of certain aspects of a problem, depending upon the terms of reference of the research. The temptation to spread an approved "gospel," whether he deeply believes in it or not, is particularly strong if the researcher (or the institution to which he belongs) is more or less entirely dependent on the financial support of the agency that commissioned the project. Instead of confining his role to research, and studying the problem with the aim of increasing understanding, he easily becomes an evangelist and not a healthy "skeptic" to whom the slogan of the day or the policy statement is a hypothesis, which, at best, lends itself to scientific inquiry.

REFERENCES

- Boalt, G. *Skolutbildning och skolresultat för barn ur olika samhällsgrupper i Stockholm*. Stockholm: Norstedts, 1947.
- Bromsjö, B. *Samhällskunskap som skolämne*. Stockholm: Scandinavian University Books, 1965. (Summary in English.)
- Conant, J. B. *The American high school today*. New York: McGraw-Hill, 1959.
- Dahllöf, U. *Kursplaneundersökningar i matematik och modersmålet* (Rep. No. 15). Stockholm: Statens Offentliga Utredningar, 1960.
- Dahllöf, U. *Kraven på gymnasiet: Undersökningar vid universitet och högskolor i förvaltning och näringsliv* (Rep. No. 22). Stockholm: Statens Offentliga Utredningar, 1963.
- Dahrendorf, R. *Arbeiterkinder an deutschen Universitäten*. Tübingen: Mohr (Paul Siebeck), 1965.
- Early leaving*. London: Her Majesty's Stationery Office, 1954.
- Ekman, G. Om uppskattningen av begåvningsreservens storlek. *Pedagogisk Tidskrift*, 1949, 6, 125-150.
- Ekman, G. Skolformer och begåvningsfördelning. *Pedagogisk Tidskrift*, 1951, 8, 15-37.
- Elmgren, J. *School and psychology*. Stockholm: Statens Offentliga Utredningar, 1952.
- Härnqvist, K. *Reserverna för högre utbildning* (Rep. No. 11). Stockholm: Statens Offentliga Utredningar, 1958.
- Härnqvist, K. The estimation of the reserves of ability. In A. H. Halsey (Ed.), *Ability and educational opportunity*. Paris: Organisation for Economic Co-operation and Development, 1959.
- Härnqvist, K., & Graham, A. *Vägen genom gymnasiet: Elevernas syn på valsituationer och studieformer* (Rep. No. 15). Stockholm: Statens Offentliga Utredningar, 1963.
- Husén, T. Intelligenskrav på olika skolutbildningsstadier. *Skola och samhälle*, 1946, 27, 1-23.
- Husén, T. Begåvningsurvalet och de högre skolorna. *Folkskolan-Svensk Lärartidning*, 1947, 1, 124-137.

- Husén, T. *Begåvning och miljö*. Stockholm: Almqvist & Wiksell, 1948.
- Husén, T. *Testresultatens prognosvärde*. Stockholm: Almqvist & Wiksell, 1950.
- Husén, T. Loss of talent in selective school systems. *Comparative Education Review*, 1960, 4, 70-74.
- Husén, T. *Problems of differentiation in Swedish compulsory schooling*. Stockholm: Scandinavian University Books, 1962.
- Husén, T. The relation between selectivity and social class in secondary education. *International Journal of Educational Sciences*, 1966, 1, 17-27.
- Husén, T. *Talent, opportunity and career*. Stockholm: Almqvist & Wiksell, 1969.
- Husén, T. *Social influences on educational attainment*. Paris: Organisation for Economic Co-operation and Development, 1975. (Also published in French and German.)
- Husén, T., & Dahllöf, U. *Matematik och modersmålet i skola och yrkesliv*. Stockholm: Studieförbundet Näringsliv och samhälle, 1960.
- Jackson, B. *Streaming: An educational system in miniature*. London: Routledge & Kegan Paul, 1965.
- Johanson, E. *Kursplaneundersökningar i fysik och kemi*. Stockholm: Almqvist & Wiksell, 1962.
- Malmquist, E. *Barnens kunskaper och färdigheter vid skolgångens början*. Stockholm: National Board of Education, 1961.
- Marklund, S. *Skolklassens storlek och struktur*. Stockholm: Almqvist & Wiksell, 1962.
- Miller, T. W. G. *Values in the comprehensive school* (Educational Monographs). Birmingham, England: University of Birmingham, 1961.
- Moberg, S. *Vem blev student och vad blev studenten?* Lund: Gleerups, 1951.
- Moberg, S., & Quensel, C. E. *1949 års universitetsberedning*. Stockholm: Statens Offentliga Utredningar, 1949.
- National Board of Education. *Försöksverksamhet med nioårig enhetsskola*. Stockholm: Author, 1959.
- Neymark, E. *Ungdomens vägledning till utbildning och yrke* (Rep. No. 43). Stockholm: Statens Offentliga Utredningar, 1945.
- Neymark, E. *Universitetens och högskolornas rekryteringsreserv*. In *Vidgat tillträde till högre studier*. Stockholm: Statens Offentliga Utredningar, 1952.
- Neymark, E. *Selektiv rörlighet*. Stockholm: Personaladministrativa rådet, 1961.
- Orring, J. *Flyttning, kvarsittning och utkuggning i högre skolor i relation till folkskolans betygssättning* (Rep. No. 35). Stockholm: Statens Offentliga Utredningar, 1959.
- Paulston, R. G. *Educational change in Sweden: Planning and accepting the comprehensive school reforms*. New York: Teachers College Press, Columbia University, 1968.
- Prop. No. 70. *Kungl. Maj:ts Proposition 1950-70 angående riktlinjer för det svenska skolväsendets utveckling*. Stockholm: Riksdagen, 1950.
- Sjöstrand, W. *Skolberedningen och differentieringsfrågan*. Stockholm: Natur och Kultur, 1961.
- Statens Offentliga Utredningar. *Den psykologiska forskningens nuvarande ståndpunkt i fråga om den psykiska utvecklingen hos barn och ungdom m.m.* (Rep. No. 19). Stockholm: Author, 1943.
- Statens Offentliga Utredningar. *Sambandet mellan folkskola och den högre skola* (Rep. No. 21). Stockholm: Author, 1944.
- Statens Offentliga Utredningar. *Betänkande och förslag angående studentsociala stöddåtgärder avgivet av 1946 års utredning om den högre undervisningens demokratisering (studentsociala utredningen)* (Rep. No. 42). Stockholm: Author, 1948. (a)

- Statens Offentliga Utredningar. *1946 års skolkommissions betänkande med förslag till riktlinjer för det svenska skolväsendets utveckling* (Rep. No. 27). Stockholm: Author, 1948. (b)
- Statens Offentliga Utredningar. *Grundskolan: Betänkande avgivet av 1957 års skolberedning* (Rep. No. 30). Stockholm: Author, 1961.
- Svensson, N. E. *Ability grouping and scholastic achievement*. Stockholm: Almqvist & Wiksell, 1962.
- Svensson, N. E., & Björklund, E. Educational research and development in Sweden. *Educational Sociology*, 1966, **39**, 145-161.
- Undeutsch, U. Auslese für und durch die höhere Schule. In *Bericht über den 22. Kongress der deutschen Gesellschaft für Psychologie*. Göttingen: Hogräfe, 1955.
- Wolfe, D. *America's resources of specialized talent*. New York: Harper, 1954.
- Yates, A. (Ed.). *Grouping in education*. Stockholm: Almqvist & Wiksell, 1966.

Chapter 9

EDUCATION, SOCIAL SCIENCE, AND THE LAW: THE RODRIGUEZ CASE AND SCHOOL FINANCE REFORM

by

MARVIN BRESSLER

Princeton University

and

JOEL S. BERKE

Education Policy Research Institute

Educational Testing Service

DISPUTES over how best to generate and allocate funds for tax-supported schools are as old as public education. They involve the interplay of philosophical concepts like equality and quality in education with the most direct of vested interests in tax bills and the availability of educational services. In state legislatures and executive chambers, victors in such contests have traditionally translated their success into law. Since the mid-1960s, however, those who lose in the traditional forums have begun to repair to the courts, there to renew the challenge. Since judicial verdicts rest both on fact and on legal principle, the protagonists often enlist the aid of social scientists, whose theories and research provide arguments for all parties to the litigation.

One aspect of this recent "judicialization" of educational policy conflict is the emergence of a fascinating and significant triangular relationship between education, law, and social science. No court proceedings are more instructive in this regard than *San Antonio v. Rodriguez* (411 U.S. 1, 1973), a case of immense social importance which could have required the

We are indebted to the National Lawyers' Committee for Civil Rights Under Law for opening to us their extensive files on all stages of the *Rodriguez* litigation.

transformation of arrangements for raising and distributing educational revenues in every state except Hawaii. As a dramatic test of social science as an instrument of advocacy in educational policymaking, the *Rodriguez* case illustrates the interaction between a scholarly tradition, empirical evidence, and legal reasoning. This chapter describes those interrelationships and attempts to trace their implications for broader issues of social science and policy development in school finance.

1. THE RODRIGUEZ CASE: A SYNOPSIS

In San Antonio, Texas, the cost and quality of public education during the late 1960s varied dramatically among the 12 independent school districts that served the city and its immediate environs, largely because of differences in the taxable wealth of each district. In Alamo Heights, the highest spending and wealthiest district, more than twice the state and local funds were available for the schooling of each pupil (\$558) than in the nearby Edgewood District (\$248 per pupil). Teachers in Alamo Heights were better paid, better educated, and proportionately more numerous.

While state aid gave each district about the same amount of money per pupil—\$225 in Alamo Heights, \$222 in Edgewood—the richer district raised an additional \$333 per pupil from local property taxes, while the poorer district raised only \$26 per pupil. The irony of the school-finance system, however, was that Alamo Heights raised those funds with only a .68% equalized property-tax effort, while the residents of Edgewood, overwhelmingly poor and of minority ethnic status, would have had to tax themselves at a rate of 5.76% to raise the same \$333 per pupil—in other words, at a rate nearly nine times that of Alamo Heights.

In the face of what they took to be state-sanctioned discrimination, a group of Edgewood parents retained an attorney and sued state and county officials. In litigation they presented evidence that the plight of their district was not an isolated example but a manifestation of a systematic pattern throughout the state. On December 23, 1971, a three-judge federal district court ruled unanimously for the plaintiffs. Characterizing the Texas system of school finance as one which "subsidized the rich at the expense of the poor" (*Rodriguez v. San Antonio*, 337 F. Supp. 280–282 [WD Tex. 1971]), the court struck down the state's school-funding laws as a denial of the equal protection of the laws guaranteed by the Fourteenth Amendment to the federal Constitution. The state legislature was ordered to pass new legislation that would no longer base the quality of a child's education upon the wealth of the local school district.

Before the legislature could act, however, the state of Texas appealed the decision to the United States Supreme Court, which, after granting jurisdiction, receiving briefs, and hearing oral argument, handed down its

decision on March 21, 1973. By the narrow margin of five to four, the high court overturned the district court decision. The majority opinion was written by Justice Powell, joined by the three other Nixon appointees, Blackmun, Burger, and Rehnquist, and by Justice Stewart. Their opinion acknowledged that the Texas system of raising and distributing revenues for education needed reform, but they held that the system did not violate traditional equal-protection standards of constitutionality. The plaintiffs were obliged to settle for the cold comfort of spirited dissents from the "liberal" wing of the Court—Justices Marshall, Douglas, Brennan, and White—for the decision foreclosed the possibility of using the federal equal-protection clause to eliminate the fiscal aspects of inequality of educational opportunity in Texas.

While the Supreme Court's decision applied directly only in the state of Texas, it had major implications for public education in the rest of the nation. The Texas approach to raising and distributing educational revenues differed only in degree, procedure, and detail from the school-funding systems employed by most other states. Moreover, the lower court opinion which had been reversed was nearly identical to decisions such as *Serrano v. Priest* (5 Cal. 3rd 584 [1971]), which had invalidated the school-finance system of California, as well as its counterpart cases in Minnesota, Kansas, Michigan, and Arizona which had been decided during the preceding 18 months. Like *Rodriguez*, these cases had been litigated under a legal theory carefully nurtured since 1965 by legal scholars, civil rights lawyers, and experts on school finance. Although these cases were not before the Supreme Court, the *Rodriguez* decision cast doubt on their authority. Furthermore, at the time of the Supreme Court decision in *Rodriguez*, pending suits in more than twenty states rested on the same theory; thereafter such cases were abandoned or substantially revised. Thus the impact of the *Rodriguez* decision extended far beyond Texas; it ended the use of federal constitutional litigation as a vehicle for school-finance reform.

2. THE STUDY OF SCHOOL FINANCE: ADVOCACY AND ECLECTICISM¹

Although the development of the legal challenge to school-funding systems dated only from the mid-1960s, the tradition of research geared to bringing change in public policy has been central to the study of school finance since the work of Elwood P. Cubberley (1906) at the turn of the 20th century. The field of school finance has been distinguished by two major tendencies: First, the field has been oriented toward advocacy

¹ For a fuller discussion of research on school finance, from which the current discussion borrows extensively, see Berke and Moskowitz (1977).

rather than "neutral" scholarship and, second, it has been willing to borrow from a variety of disciplines and traditions.

Its earliest practitioners analyzed fiscal aspects of the provision of public education, formulated legislative approaches to funding problems, and provided advisors and consultants to interest groups and public agencies. In the process, scholars have developed normative standards of equity in taxation and equality of opportunity in the provision of services and have focused empirical research on the degree to which existing laws meet or depart from such standards. While advocacy has often taken the form of arguments for particular approaches to educational financing, in recent years it has also emphasized the use of empirical analysis and computer simulation techniques to assist policymakers in assessing more accurately the probable impact of a variety of alternative modes of financing schools.

The second prominent characteristic of research in school finance is its eclectic quality. From an early base in educational administration, it has reached out to incorporate concepts and analytic methods from other fields. Because of the magnitude of the costs of public education, prominent students of public finance have turned their attention from time to time to the study of the processes through which revenues for public elementary and secondary schools are raised and distributed, and these contributions from economics have posed new questions and redirected research interests within the field of school finance. Most recently, the concerns of the legal profession have had their impact. School finance has adapted its methods of analysis to respond to the requirements of equal-protection litigation and, in so doing, has appropriated some of the language of the law.

2.1 *The Early Tradition of Scholarship in School Finance*

The work of Elwood Cubberley is generally taken as the first identifiable research in school finance. His studies in several states convinced him of the marked variation in the quality of school programs available throughout most states, and he developed new legislative formulas intended to enhance the quality of educational offerings, such as making kindergarten and high schools available to poorer districts, which were frequently located in small towns and rural areas. At the same time, he thought it important to encourage greater fiscal effort for education in those districts—usually the cities which sought to maintain higher than average standards in curriculum offerings and facilities. Cubberley properly identified variations in local tax bases as the chief source of inequalities in basic programs and vividly illustrated the plight of property-poor communities that taxed themselves at high rates but realized only limited revenues. His solutions, however, were too modest for their intended purposes, and while the formulas he developed provided incentives for offering a longer school year, a more plentiful supply

of teachers, and a more extensive curriculum, they failed to overcome, and may even have unintentionally exacerbated, the problems of poorer school systems. But Cubberley's work was to influence state-aid formulas for decades to come, and his emphasis on tailoring the form of the fiscal instrument to serve educational purposes continues to be central to the field he initiated.

The next major step in the development of school-funding systems was made in the early 1920s by George D. Strayer and Robert M. Haig (1923), who gave primary emphasis to equalizing the taxable-resource disparities between school districts. Their contribution, the *foundation plan*, provides that the state defines the costs of a satisfactory minimum school program, requires that localities make a predetermined tax effort to fund such a program, and uses the state grant-in-aid mechanism to make up the difference, if any, between the yield of the mandated local tax rate and the amount of the minimum program. Strayer and Haig assumed that the required local effort would be set at the level of the tax rate necessary for the richest district to raise the amount established by the "foundation" program, but in practice states tended both to guarantee the yield of a lower rate and to set the "foundation" level well below the average statewide expenditures. In addition, to encourage innovation in "lighthouse" districts, the plan permitted localities to augment the foundation program by raising as much additional revenue as they wished. The result of this local leeway was to permit rich districts to realize much higher educational spending levels with lower tax efforts, in much the same manner as earlier flat grant systems. Nonetheless, the foundation program remains the most widespread basic structure of state finance.

A competing plan of equalization known as the *percentage equalizing grant* was developed during the same period (Updegraff & King, 1922). It permitted localities to select their own level of school support, with the state sharing in the costs under a formula inversely proportional to the size of the local property-tax base. In theory, this approach represented another significant advance in equalizing tax burdens and revenues available to rich and poor districts, thus reducing discrimination against children who resided in the latter. In practice, the few states that adopted percentage equalizing formulas limited the level of expenditures subject to equalization or simply failed to provide sufficient revenue at the state level to overcome the basic resource disparities between districts.

A fourth approach to the financing of elementary and secondary education was proposed at the end of the 1920s by Henry Morrison (1930). Based on the assumption that the purpose of the public schools was to socialize young people into responsible citizenship, Morrison felt that compensatory spending would be necessary for the children of city and rural slums. He believed that such expenditure patterns would be achieved only if the state assumed full responsibility for the costs and distribution of educational revenues. All other approaches would permit

wealthy local districts to spend at will, thus nullifying or reducing the equalizing impact of compensatory funds. High spending in wealthy districts served only what Morrison considered purely private—as opposed to social or civic—educational purposes. *Full state assumption* was actually adopted in the late 1960s by Hawaii, and although no other state has followed suit, it has had lasting impact on reformist thinking, frequently appearing as a goal of state study commissions and as an influence on the current trend toward increasing the state share of school revenues.

In the following years, the work of these early school-finance specialists, as refined by successors such as Paul Mort, Arvid Burke, R. L. Johns, Erick Lindman, and others, provided the models for legislation in virtually all states. However, the actual legislation was typically enacted with significant limitations: floors to ensure that all districts, no matter how wealthy, received some aid; ceilings to limit the aggregate cost of the percentage equalizing programs; and foundation levels set far below average expenditure levels.

By the late 1940s and early 1950s, another shortcoming of these programs became apparent. Designed for a time when cities were fiscally healthy and outlying areas were impoverished, equalization formulas were intended to redistribute funds from cities to small towns and rural areas. Such formulas continue to discriminate against large cities, even though the most populous urban centers have since experienced a marked decline in their comparative fiscal health. While this phenomenon has multiple causes, the chief problem arises from the use of property value per pupil to measure fiscal capacity. Many large cities appeared “rich” by the simple measure of property value per pupil, and therefore their state aid tended to fall below the average. More sophisticated indicators of taxpaying ability would have shown many cities to be in a far more disadvantaged position (Sacks, 1972).

Some suburban and small-town districts, those lacking a commercial and industrial tax base, also fared badly under traditional equalization formulas. Those localities often had such low property-tax rolls that their state aid was too small to overcome this disadvantage in local revenues. Even higher than average local tax effort often failed to yield revenues as large as the state average. In short, the various equalization formulas generated insufficient funds to overcome disparities and permitted gross inequalities of expenditure among districts.

2.2 *The Influence of Economics on School Finance*

During the 1960s a new generation of economists became interested in school finance as part of research in the larger context of the public economy. Charles Benson (1961, 1963) and Jesse Burkhead (1964) were among the first economists to specialize in this area. Focusing less on the specific aid mechanism and more on theoretical and empirical inquiry into

systems of taxation and intergovernmental aid, economists made important contributions to at least five areas.

Richard Netzer's *Economics of the Property Tax* (1966) contributed valuable insights into the ability of certain classes of property owners to shift the incidence of taxation to others. Despite recent reexamination and alternative theories on this question (e.g., Aaron, 1975), Netzer's discussion of the regressivity of property tax remains a staple of school-finance reform. During this same period, economists have also examined the effects of different proportions of types of property (residential, nonresidential, etc.) on the fiscal behavior of local jurisdictions (Ladd, 1975).

A second group of influential economic studies has documented economic and demographic shifts since World War II (U.S. Advisory Commission on Intergovernmental Relations, 1967). Studies by Harvey Brazer, Alan Campbell and Seymour Sacks (Campbell & Sacks, 1967), and others described shifts in economic activities and population patterns that left the nation's largest cities with stagnant tax bases and costly social needs, while suburban areas were drawing new businesses and higher income populations with relatively little need for public health, welfare, law enforcement, and other public services. The demonstrated impact of these phenomena on the fiscal capacity of different types of jurisdictions has led to the inclusion, in recent years, of "urban factors" and income measures as adjustments to the property-value standard in school-finance formulas.

In addition to taxation and fiscal capacity, a third set of issues, those of distribution, has attracted economists' concern. One prominent scholar, Werner Hirsch (1970), posed these questions: Does aid increase expenditures from sources available to the recipient government? Does aid distort the allocation of resources between the recipient government and the private sector? Does aid distort the allocation of resources among services offered by recipient governments? Pursued by Hirsch and others, this line of inquiry has led to an examination of the components of state aid to local school districts. What are the implications of a flat grant for local fiscal behavior? What results can be anticipated when a state pays all transportation and teacher pension costs? As empirical studies were undertaken to examine a local district's response to changes in the school-finance laws, the economic principles of income and substitution effects emerged as important guides for inquiry. For example, such questions arise as whether additional state aid will be used by local school districts to increase school expenditures, to offset revenues derived from local property taxes, or to achieve predictable combinations of both sets of activities.

Fourth, economists have also undertaken to describe the determinants of public expenditure behavior (Bahl, 1968; Burkhead & Miner, 1971). In most cases the principal components of public expenditure models were cross-sectional data, expenditure levels as the dependent variable, and

multivariate statistical techniques as the methodological tool. The findings can be summarized as establishing the relationship between property valuation and socioeconomic characteristics on a district's expenditure level. In the terms employed in these studies, a significant amount of the variation in local district expenditure patterns was explained by these major independent variables (James, Kelly, & Garms, 1966). The linkage between wealth and expenditures—and subsequent analysis relating expenditures to educational offerings—has been instrumental in establishing the unconstitutionality of numerous state educational aid systems.

Finally, economists (Burkhead, Fox, & Holland, 1967; Thomas, 1971) adapted analytic methods used in examining the relationships between factors of production and industrial output to measure and trace the influence of factors that affect the output of public education. Despite much continuing study, the complexities of the educational process have not readily yielded to the methods of production function analysis; in most such studies, school factors seemed to have little impact compared with socioeconomic characteristics such as parental income (Coleman, Campbell, Hobson, McPartland, Mood, Weinfield, & York, 1966). Some recent research, however, which has utilized more precise data or more intensive analytic methods, has suggested that when carefully described process factors can be more precisely matched over time with particular students, significant differentiation between the effects of resource inputs can be found (Wiley & Harnischfeger, 1974; Sommers & Wolf, 1975). It is not yet clear, however, whether social scientists can develop sufficiently refined methods to isolate the impact of the complex factors that affect educational performance.

2.3 *Legal Scholarship and the Reform of School Finance*

The effect of economics on school finance was to open new areas for inquiry, to provide more powerful methods of analysis, and to emphasize the ultimate impact, rather than the internal detail, of aid systems. But beginning about five years after this tendency had appeared in the field, another scholarly orientation also began to influence research in school finance: Legal strategists suggested a means for closing the gaps between theoretical policy solutions and imperfect legislative implementation. Given the orientation of school-finance scholars toward advocacy, the overtures of law professors and civil rights litigators were warmly welcomed. The leading names are those of Arthur Wise, who first publicly suggested a constitutional theory that would employ the equal-protection clause to invalidate inequitable state finance laws, and John C. Coons, who perfected the litigation strategy that was to prevail in the *Serrano* case and in most other cases both before and after *Rodriguez*.

In a 1965 article in the *Administrator's Notebook* and at greater length in *Rich Schools, Poor Schools* (1968), Arthur Wise argued that major constitutional precedents, were they applied to educational finance patterns,

would result in the finding that existing state systems of educational finance were unconstitutional. *Brown v. Board of Education* (347 U.S. 483 [1954]) had established that education was of fundamental constitutional importance and must be made available to all children on equal terms. *Griffin v. Illinois* (351 U.S. 12 [1956]) and *Douglas v. California* (372 U.S. 353 [1963]) had declared that a defendant's right to an effective defense could not be invidiously conditioned by his poverty and *Baker v. Carr* (369 U.S. 186 [1962]) had declared that the right to vote should not be made to vary in its effectiveness by the geographical residence of a voter. Wise suggested that, taken together, these cases would establish that a state denied its school children the equal protection of the laws guaranteed by the Fourteenth Amendment when it permitted the quality of education within its borders to vary because of geography (the particular district in which a child lived) or poverty (the variations in local real-property tax bases). In place of these inappropriate criteria for allocating educational resources, he suggested that only a child's need for educational services or equal educational opportunity was appropriate as the state's distributional criterion.

The first case to test this novel theory was *McInnis v. Shapiro*, in 1968 (293 F. Supp. 327 [ND Ill]), and the judges were not convinced. Dismissing the suit for failure to state appropriate constitutional grounds for relief, a federal district court in Illinois held that while "the inequalities of the existing arrangements are readily apparent . . . the allocation of public revenues is a basic policy decision more appropriately handled by a legislature than a court." The court was not convinced that Illinois's basic funding system, despite the disparities in spending it created, was entirely irrational, since its emphasis on decentralization in educational decision-making was a legitimate legislative goal. But the court's judgment was also influenced by its fear of the nature of the relief requested by the plaintiffs: a court-ordered funding system that would allocate educational revenues on the basis of the "needs" of pupils in various districts.

Even if the Fourteenth Amendment required that expenditures be made only on the basis of pupils' educational needs, this controversy would be non-justiciable. . . . There are no "discoverable and manageable standards" by which a court can determine when the Constitution is satisfied and when it is violated. (*McInnis v. Shapiro*, 293 F. Supp. 327 [ND Ill])

Although *McInnis* (affd. per curiam sub nom. *McInnis v. Ogilvie*, 394 U.S. 322 [1969]) was affirmed without opinion by the U.S. Supreme Court a year later, lawyers, school-finance scholars, and civil rights activists were not convinced that their cause was doomed. Numerous conferences, articles, and plea drafts honed the legal strategies that would later be successful. The scholars who provided the successful legal formula were John C. Coons, then a professor of law at Northwestern Law School, and two of his students, William Clune, III, and Stephen Sugarman. In an influential

law review article (Coons, Clune, & Sugarman, 1969) and a subsequent book (Coons, Clune, & Sugarman, 1970), they rejected the earlier argument that the equal-protection clause required the state to distribute educational resources in proportion to educational need. Instead, the only order they requested of the courts was that states should finance public education in a manner that does not make it easier for wealthy communities to provide a better education than poor school districts can offer. This theory came to be known as the *test of fiscal neutrality*: Only the wealth of the state as a whole, not of its individual school districts, could condition the money available for a child's education. Fiscal neutrality became the explicit rationale of the California case, *Serrano v. Priest*, in which the California Supreme Court ruled that school-finance systems "[should not] invidiously discriminate against the poor [by making] the quality of a child's education a function of the wealth of his parents and neighbors" (5 Cal. 3d 584).

The Wise and Coons conceptions reflected rival approaches to legal strategy—the latter required less explicit judicial policymaking—but they were also informed by different philosophical emphases. Wise's *Rich Schools, Poor Schools* (1968) was more obviously egalitarian, urging that school resources be distributed in a way that ensured at least equal treatment, and, at best, a compensatory approach designed to use public resources to overcome learning disadvantages of social and economic origin. The book *Private Wealth and Public Education* by Coons, Clune, and Sugarman (1970) emphasized the principle of "subsidiarity," the right of individual units to select different levels of educational offerings on the basis of choice unimpeded by differences in wealth. Where Wise and others were primarily concerned with equality in educational offerings, the "power-equalizing"² approach of Coons and his colleagues was

² District Power Equalizing (DPE—developed by Coons, Clune, and Sugarman) was a modification of existing equalization plans. DPE is thoroughly consistent with the "capacity equalization" tradition which distributes state aid in inverse proportion to local taxable resources. It is innovative, however, both in emphasis and in substance. DPE focuses on the effort factor and guarantees that, for any given level of tax effort, all communities will be guaranteed an equal level of spending through a combination of local and state revenues. This plan is usually presented as a table of guaranteed expenditure levels, which correspond to tax efforts, established by the state. For example,

Tax Effort	Guaranteed Level of Expenditure
10 mills	\$1000 per pupil
20 mills	\$2000 per pupil
30 mills	\$3000 per pupil

The district chooses a level of tax effort. If it is unable to generate revenues equal to the guaranteed level of expenditures, the state makes up the difference.

Another feature distinguishing this plan from traditional formulas is the possibility of "recapture" of local revenues. Local districts that raise an amount per pupil in excess of the state-guaranteed amount would have to pay back the excess to the state for redistribution to poorer schools (i.e., those with less valuation per pupil). No earlier school-equalization formulas included such recapture. At least four states (Utah, Maine, Montana, and Wisconsin) have recently enacted district-power-equalizing provisions with recapture (Goertz & Moskowitz, 1976).

primarily aimed at providing tax-rate equity so that any level of tax effort would produce equal educational resources for all school districts. Both of these theories were prominent in the *Rodriguez* litigation.

3. SAN ANTONIO V. RODRIGUEZ: LAW AND SOCIAL SCIENCE

The case of *San Antonio v. Rodriguez*,³ as noted above, was initiated by parents of school children in the Edgewood School District, one of several independent school districts located within the City of San Antonio. The suit was brought in the federal courts because it involved a claim under the U.S. Constitution. The plaintiffs requested, and were granted, a special three-judge district court because the case involved an important issue of constitutional law. The value of the three-judge court is that it speeds the process of subsequent appeal, since appeals from a three-judge court go directly to the U.S. Supreme Court, without the intermediate step of being heard first by a U.S. court of appeals.

Given the final outcome of the litigation, the use of the federal courts was probably a strategic blunder. The national network of civil rights attorneys and law school professors interested in school-finance suits has wanted to avoid appeals to the U.S. Supreme Court as long as possible out of fear of what its conservative-majority high court might do with an issue of such far-reaching social importance. However, the San Antonio attorney who brought the suit was not affiliated with the national network and followed his own sense of where he was likely to get a more sympathetic hearing in Texas.

The focus of the argument was the established legal framework of equal protection, which was not challenged in the course of the *Rodriguez* litigation by any of the contesting parties. As summarized by Justice Powell, it required the following analytic steps:

We must decide, first, whether the Texas system of financing public education operates to the disadvantage of some suspect class or impinges upon a fundamental right explicitly or implicitly protected by the Constitution, thereby requiring strict judicial scrutiny. If so, the judgment of the District Court should be affirmed. If not, the Texas scheme must still be examined to

³ A note on terminology. In the district court, the case was identified by the first-named plaintiff in the pleadings (Demetrio P. Rodriguez) and the first-named defendant (the San Antonio School District) and was docketed as *Rodriguez v. San Antonio*. On appeal, the names were reversed since it was the defendants who, having lost in the district court, brought the case to the U.S. Supreme Court, and the case at that level was designated *San Antonio v. Rodriguez*, although the less formal appellations—the *Rodriguez* case, or simply *Rodriguez*—continue to be used. The parties at the Supreme Court level are generally referred to in the opinions as the Appellants (the defendants in the original proceedings) and the Appellees (the plaintiffs in the lower courts).

While the case takes its name from the San Antonio School District, in fact, the State of Texas was the principal defendant and appellant throughout the litigation.

determine whether it rationally furthers some legitimate, articulated state purpose and therefore does not constitute an invidious discrimination in violation of the Equal Protection Clause of the Fourteenth Amendment. (*San Antonio v. Rodriguez*, 411 U.S. 1, 17 [1973])

Each of the key terms in the preceding passage—"strict scrutiny," "suspect class," "fundamental right," "rational" or, alternatively, "compelling," "state purpose"—requires further elaboration.

The law is ordinarily guided by the doctrine of "presumptive constitutionality." The courts generally assume that legislatures are sensitive to constitutional principles, constrained by public opinion, knowledgeable about local conditions and events, and thus may be trusted to pass laws which, if not always wise, are nevertheless defensible by the standards of democratic politics. This assumption prevails for the most part even in those numerous instances when a legislature discriminates for legal purposes between various segments of the population. Thus, for example, all states withhold the privilege of driving a motor vehicle from persons below a certain age; no court has seen fit to surround juveniles thus denied with the mantle of constitutional protection. Concern for public safety fortified by reasonable expectations about the nature of youthful behavior presumably justifies such legally imposed, invidious distinctions between age categories.

The courts, however, do not always take this indulgent view of legislative intent and action. According to honored judicial precedent, some forms of discrimination, notably those based on race, nationality, religion, wealth, alienage, and sex, are regarded as "suspect" and as such invite "strict scrutiny." When such criteria mandate unequal treatment of groups or categories, a state may be required to show that it has not violated the "fundamental rights" of the designated population and that the statute is responsive to a "compelling" rather than simply to a "rational" state interest.

The concept of fundamental right is not synonymous with social importance or distributive justice; more is required than proof that various forms of systematic deprivation suffered by a group injure its life chances, prestige, or power. To qualify as "fundamental" a right must be explicitly or implicitly guaranteed by the Constitution. Thus, unequal access to housing or medical care may be a personal and social misfortune, but since neither is judged to be constitutionally guaranteed, relief must be sought through the political rather than the judicial process. A poll tax, however, which is arguably no more prejudicial to individual and social welfare than unequal access to housing or medical care, violates the intent of the First Amendment and therefore comes within the purview of the courts.

Even when the basis for discrimination is "suspect" and a "fundamental right" is abridged, a state may reward some groups and penalize others if

it can establish that a statute rests on a "compelling state interest," that is, it serves paramount values which have even greater social claims than the doctrine of equality under the law. The questionable incarceration of Japanese-American citizens in detention camps during World War II is perhaps the most dramatic example of the traditionally impermissible violation of the fundamental rights of a victimized population in the name of a compelling interest, that is, military necessity. Such legislation could not be defended merely on the grounds that it was "reasonable" or "rational." The term "compelling," then, implies that the interest of the total collectivity may supersede the rights of its constituent elements but only under very rare circumstances and when other courses of action are unavailable or severely restricted.

3.1 *The Judicial Resolution of Educational Issues*

In applying the general principles of equal-protection analysis to the *Rodriguez* case, the Supreme Court was obliged to identify seven major educational issues bearing on the case.

1. May a state distribute its educational resources unequally on the basis of the unequal wealth of its recipients?
2. Shall the poor children or alternatively children living in tax-poor school districts be defined as the class requiring constitutional protection?
3. Is education "fundamental" or merely important to the welfare of the nation and its citizens?
4. Shall educational quality be defined by the resources made available to children or by the outcomes that are demonstrably attributable to schooling?
5. Shall the quality of education, however defined, be conceived as equal where it is fashioned to the individual needs of the child, is the same as is offered to the most privileged, or satisfies the requirements of a specified adequate standard?
6. What are the positive and negative consequences of a school finance system that relies on the local property tax for the general social and educational goals of the state?
7. Are there alternative feasible methods of financing schools to remedy the defects of the present arrangements yet retain their desirable features?

May educational resources be distributed unequally? The basic nature of the Texas system of school financing, which was similar to those in most other states, was not in dispute by the litigants:

The financing of public elementary and secondary schools in Texas is a product of state and local participation. Almost half of the revenues are derived from a largely state-funded program designed to provide a basic minimum educational offering in every school. Each district supplements state aid through an ad valorem tax on property within its jurisdiction. (411 U.S. 1)

With some minor variations, it was a traditional foundation formula with the usual problems of an inadequate foundation level and unlimited local leeway. The principal empirical evidence that was developed on

behalf of *Rodriguez* was intended to demonstrate that this method of raising and distributing educational funds, with its heavy reliance on the unequalized local property tax, was inconsistent with the principle adopted in *Serrano* and in the subsequent cases that had followed its lead: "This funding scheme invidiously discriminates against the poor because it makes the quality of a child's education a function of the wealth of his parents and neighbors" (5 Cal. 3d 584 [1971]). To the three judges of the federal district court in Texas, the evidence had been convincing. Their unanimous opinion declaring the Texas system unconstitutional included this comment on the discriminatory effect of local property taxation on the provision of public education:

Within this ad valorem taxation system lies the defect which plaintiffs challenge. This system assumes that the value of property within the various districts will be sufficiently equal to sustain comparable expenditures from one district to another. It makes education a function of the local property tax base. The adverse effects of this erroneous assumption have been vividly demonstrated at trial through the testimonial exhibits adduced by plaintiffs. In this connection, a survey of 110 school districts throughout Texas demonstrated that, while the ten districts with a market value of taxable property per pupil above \$100,000 enjoyed an equalized tax rate per \$100 of only thirty-one cents, the poorest four districts, with less than \$10,000 in property per pupil, were burdened with a rate of seventy cents. Nevertheless, the low rate of the rich districts yielded \$585 per pupil, while the high rate of the poor districts raised only \$60 per pupil. As might be expected, those districts most rich in property also have the highest median family income and the lowest percentage of minority pupils, while the poor property districts are poor in income and predominantly minority in composition.

Nor does State financial assistance serve to equalize these great disparities. Funds provided from the combined local-state system of financing in 1967-1968 ranged from \$231 per pupil in Edgewood to \$543 per pupil in Alamo Heights. There was expert testimony to the effect that the current system tends to subsidize the rich at the expense of the poor, rather than the other way around. Any mild equalizing effects that state aid may have do not benefit the poorest districts.

For poor school districts educational financing in Texas is, thus, a tax more, spend less system. The constitutional and statutory framework employed by the State in providing education draws distinction between groups of citizens depending upon the wealth of the district in which they live. (*Rodriguez v. San Antonio*; 337 F. Supp. 280, 282-283 [WD Tex. 1971])

The principal financial data relied upon in the case were provided by a study of Texas school finance directed by Joel S. Berke, then of the Syracuse University Research Corporation. Based on a stratified random sample of 110 of Texas' 1,183 districts, the study revealed a linear relationship between the amount of a district's taxable property and its expenditures per pupil for public education. A similar correlation between median family income and per-pupil expenditures was also obtained for districts at the extremes of the distribution but was not confirmed for the

districts at intermediate levels of wealth. Although the majority noted that the methodology that generated these data had been questioned in a law review article (Goldstein, 1972, n. 67), Justice Powell concluded that "insofar as any of these correlations is relevant to the Constitutional thesis presented in this case we may accept its basic thrust" (411 U.S. 1, 15-16, n. 38).

The data also showed that districts with low per-pupil school expenditures were not derelict in meeting their fiscal responsibilities. On the contrary, there was an orderly inverse relationship between the value of taxable property in a district and its tax effort, so that the equalized tax rates for the poorest districts were more than twice as high as those in the wealthiest localities. Moreover, the data also revealed that the allocation formula used by the state granted approximately equal funds to the poorer and richer districts, thus permitting greater expenditures to be made in the richer districts from their substantial local taxable resources.⁴

This description of the fiscal bias in the Texas school-finance system was emphasized by Justice Marshall in his dissenting opinion in the Supreme Court. After referring to the contrasts between the Edgewood and Alamo Heights districts, he observed:

Edgewood . . . raised only \$26.00 per pupil through its local property tax, whereas Alamo Heights was able to raise \$333 per pupil. Since the funds received [from the state] through the Minimum Foundation School Program are to be used only for minimum professional salaries, transportation costs, and operating expenses, it is not hard to see the lack of local choice—with respect to higher teacher salaries to attract more and better teachers, physical facilities, library books, and facilities, special courses, or participation in special state and federal matching funds programs—under which a property poor district such as Edgewood is forced to labor. In fact, because of the difference in taxable local property wealth, Edgewood would have to tax itself almost nine times as heavily to obtain the same yield as Alamo Heights. (411 U.S. 1, 128-129)

The issue before the Supreme Court, then, was whether a finance system based heavily on local property taxes and producing unequal per-pupil expenditures thus rested on the suspect category of wealth. In its plea before the district court, Texas denied that differentials in school spending were either immoral or illegal. Indeed, the lawyer for the state implied that "fiscal neutrality" offended the principles of democratic capitalism. According to this brief,

it would be nothing less than honest at this point to call what the Plaintiffs seek by its real name—socialized education. The whole thrust of their assertions is that regardless of the desires or wishes of the parents of school children, it will

⁴ See Appendix A for the principal conclusions of the study and the tables that were included in Supreme Court opinions. For published versions of the study see Berke, Carneyale, Morgan, and White (1972) or Berke (1974).

not be allowed for one child to receive a larger allocation of educational funds than another even if the parents in a particular school district can and are willing to pay the price. (*Rodriguez v. San Antonio*, Defendants Trial Brief, p. 7)

This line of argument did not persuade either the district court that declared the Texas school-finance system unconstitutional or, for that matter, the Supreme Court, which reversed the original verdict. The former concluded that the cases forbidding the poll tax and establishing the rights of indigents to equal treatment in criminal proceedings had confirmed that wealth was a suspect classification. It further reasoned that these precedents applied to the disparities created by local property taxes, a view with which the minority of the Supreme Court concurred. Even Justice Powell, speaking for the majority, conceded that some forms of discrimination based on wealth might require judicial intervention but concluded that "the District Court's opinion does not reflect the novelty and complexity of the constitutional questions posed by the appellees' challenge to Texas' system of school financing" (411 U.S. 1, 17). In a revealing footnote he writes,

An educational finance system might be hypothesized, however, in which the analogy to the wealth discrimination cases would be considerably closer. If elementary and secondary education were made available by the State only to those able to pay a tuition assessed against each pupil, there would be a clearly defined class of "poor" people—definable in terms of their inability to pay the prescribed sum—who would be absolutely precluded from receiving an education. That case would present a far more compelling set of circumstances for judicial assistance than the case before us today. After all, Texas has undertaken to do a good deal more than provide an education to those who can afford it. It has provided what it considers to be an adequate base education for all children and has attempted, though imperfectly, to ameliorate by state funding and by the local assessment program the disparities in local tax resources. (411 U.S. 1, 25)

Thus, even the Supreme Court's majority was prepared to acknowledge that discrimination based on wealth was constitutionally suspect in education as well as other public functions. The point in dispute was whether interdistrict disparities in per-pupil expenditures for schooling, in and of themselves, violated the Fourteenth Amendment guarantees. Of all the legal problems posed by the *Rodriguez* case, therefore, none so divided the Court as the question of defining the class that required constitutional protection. Justice Marshall⁵ held that constitutional protec-

⁵ There were, in fact, three dissenting opinions. Justice Brennan wrote a one-page dissent emphasizing his agreement with the other two dissenting opinions (411 U.S. 1, 62). Justice White, joined by Justices Douglas and Brennan, wrote a second dissent emphasizing the degree of fiscal discrimination apparent in the facts of the case and arguing that the parents and children of Edgewood were a definable class entitled to protection from a system which was without any rational basis (411 U.S. 1, 63-70). However, the major dissent was written by Justice Marshall and signed by Justice Douglas as well (411 U.S. 1, 70-137). It fully developed the entire range of legal and factual issues in the case and is the principal source of minority views discussed in this chapter. A few informed followers of Supreme Court activities, judging

tion should be extended to all children, regardless of income level, who live in property-poor school districts; the majority insisted that existing judicial precedents establishing wealth as suspect were applicable to poor individuals, not poor districts, and were not controlling when unequal educational resources were equally distributed to both rich and poor students in the same school district.

Poor children or poor districts? Of course, systems of school financing primarily based on the local property tax do not specifically require that more abundant educational resources be used for some children rather than for others: The language of such laws does not explicitly require the allocation of school funds in direct proportion to family wealth. Moreover, since some poor families reside in property-rich districts and vice versa, both rich and poor children may be said to share a common fate. Furthermore, all children in Texas enjoyed equal rights to attend tax-supported schools financed at levels deemed "adequate" by the state. This set of circumstances led the Supreme Court majority to conclude—in "the absence of any evidence that the financing system discriminates against any definable category of 'poor' people or that it results in the absolute deprivation of education—that this advantaged class is not susceptible to identification in traditional terms" (411 U.S. 1, 25). On this basis this court refused "to extend its more exacting scrutiny to review a system that allegedly discriminates against the large, diverse and amorphous class unified only by the common factor of residence in districts that happen to have less taxable wealth than other districts" (411 U.S. 1, 28).

The criterion of "common factor of residence" that seemed irrelevant to Justice Powell appeared decisive to the minority. According to the dissenting justices, judicial precedents establishing wealth as suspect required the extension of constitutional protection to all children regardless of personal family income if they live in property-poor districts. The following reasoning—as perhaps best expressed by an *amicus curiae* brief prepared by Coons, Clune, and Sugarman in support of the *Rodriguez* plaintiffs—wholly persuaded Justice Marshall.

Defendants argue at length . . . that the court's findings of a coincidence of district poverty and low family income is disputable and that the result below, therefore, is bottomed on an "unsound factual assumption." They imagine, for reasons undisclosed, that the "wealth" relevant to the constitutional rule adopted below is the income of district residents. But, as the court below held, it is district poverty which is constitutionally crucial, for this is the barrier which operates to deprive plaintiffs of their rights. (*San Antonio v. Rodriguez*, *Amicus Curiae* Brief on behalf of John Serrano, Jr., and John Anthony Serrano in support of appellees, pp. 16-17)

both from the form of the Marshall dissent and from other sources of information, suggest that it was originally intended as a majority opinion, implying that at some point in the predecision processes of the court, there may have been a fifth vote for affirming the district court decision.

It is notable that the divergent views of the court were not markedly affected by the disputes over factual evidence. Both majority and minority agreed that there was a substantial correlation between district property wealth and per-capita school expenditures. For Justice Marshall, that relationship was sufficient to demonstrate the existence of unconstitutional discrimination, but for Justice Powell it was not. For the latter to find the Texas system unconstitutional it was necessary to show an inverse relationship between personal income and school expenditures throughout the entire range of districts in the sample. For Justice Marshall, on the other hand, that relationship was constitutionally irrelevant, although he appears to have found the plaintiff's evidence on income discrimination rather convincing.⁶ Thus, it was the selection of the relevant information, not the validity of the statistical relationships, that proved decisive in resolving the issue.

From the minority's perspective, the fact that some children suffered only relative rather than absolute deprivation in access to school resources or that local taxes were augmented by state grants in no way relieved Texas of its obligation to distribute school funds equitably. Thus Justice Marshall quotes the decision in the poll tax cases which held that the "degree of discrimination was irrelevant" and adds that

the court did not order merely that those too poor to pay the tax be exempted; complete impecunity clearly was not determinative of the limits of the disadvantaged class, nor was it essential to make an equal protection claim. (411 U.S. 1, 118)

He dismisses the "equalizing effect" of the state's Minimum Foundation School Program as a "cruel irony" through which "much more state aid is being given to property rich Texas school districts on top of their already substantial local property tax revenues" (411 U.S. 1, 82) and concludes, therefore, that "Texas' financing scheme is hardly remedial legislation for which we have previously shown substantial tolerance" (411 U.S. 1, 81, n. 85).

Fundamental rights: Is education essential or merely important? Here as in other aspects of the *Rodriguez* case the same facts are capable of sustaining radically different interpretations of the Constitution. Once the majority ruled that interdistrict disparities in school expenditures were not constitutionally suspect and that strict scrutiny was not required, it could have elected to reverse the district court solely on these grounds.⁷ Such reticence, however, would have been inconsistent with the court's evident desire to issue a definitive judgment touching all aspects of

⁶ See Appendix B for excerpts from opinions discussing evidence on income discrimination.

⁷ But see the dissenting opinions of Justices Brennan (411 U.S. 1, 62) and White, Douglas, and Brennan (411 U.S. 1, 67-68) which argue that the Texas School Finance System cannot pass even the more lenient "rational basis" test of equal protection.

equal-protection analysis. In doing so, the majority adopted the familiar form of debate that initially denies an opponent's basic premises and then proceeds to confront his other arguments by recourse to "even if we were to concede . . . it still would not follow that . . ."

Justice Marshall, on the other hand, having concluded that strict judicial scrutiny was warranted, was obliged to show that children in property-poor districts had been deprived of equal opportunity to enjoy a fundamental right. A necessary, but insufficient, ingredient of any fundamental right is that it has important consequences for the individual and society. Judicial concern is obviously greater when the effects of discrimination are serious rather than trivial. That this was indeed the case was asserted by the attorneys for *Rodriguez*, conceded by the state, and affirmed by the court. The brief for the plaintiffs refers to five principal functions of education: First, "essential to the maintenance of the free enterprise system"; second, "vital to the development of the individual"; third, "universally relevant and . . . compulsory"; fourth, "a vital element in the moulding of personality"; and, fifth, "vital to the economic and political survival of the state" (*San Antonio v. Rodriguez*, brief for the Appellees, p. 9). These themes and their variations were fully elaborated in the *Rodriguez* litigation. Both the Powell and the Marshall opinions quoted approvingly the unanimous view of the court in *Brown v. Board of Education* that "in these days, it is doubtful that any child may reasonably be expected to succeed in life if he is denied the opportunity of an education" (347 U.S. 483, 493 [1954]).

It does not follow, however, that because education is desirable or even indispensable to the welfare of the nation and its citizens, it is necessarily "fundamental" for purposes of equal-protection analysis. Equal schooling is protected by the Fourteenth Amendment only if it is linked to rights that are explicitly or implicitly guaranteed elsewhere in the Constitution. Since the Constitution makes no direct mention of education, the plaintiffs were obliged to claim (and the state to deny) that equal educational opportunity is a prerequisite for the exercise of other rights that are indisputably fundamental because they rest on the specific or derived authority of the First Amendment.

The brief for the appellants well summarizes the major argument.

While, no doubt, the First Amendment protects uninformed, unpersuasive, and possibly unintelligible speech, the meaningful exercise of that right, enabling the speaker to convince and persuade, is dependent upon his ability to speak intelligently and knowledgeably. This is the essence of the free speech guarantee. It is not the act of vocalizing meaningless sounds that is the *raison d'être* for judicial protection; rather it is the process of exchanging ideas for the purpose of gaining "wisdom in action." (pp. 31-32)

Similar contentions were advanced in the *amicus curiae* brief submitted by the National Education Association and associated groups which cited the existing research on political behavior in order to show that "virtually

all studies on the subject have found a strong positive relation between educational attainment and political participation" (*San Antonio v. Rodriguez*, Motion for leave to file brief and brief *Amicus Curiae* on behalf of National Education Association et al., p. 23).

These views, which were incorporated in the minority opinion, represent a broad construction of the First Amendment, that is, free speech should be conceived as a meaningful, not merely a formal, right. The Powell opinion reflects a less activist, more orthodox view of the role of the courts in support of First Amendment rights.

We need not dispute any of these propositions. The court has long afforded zealous protection against unjustifiable governmental interference with the individual's rights to speak and to vote. Yet we have never presumed to possess either the ability or the authority to guarantee to the citizenry the most *effective* speech or the most *informed* electoral choice. That these may be desirable goals of a system of freedom of expression and of a representative form of government is not to be doubted. These are indeed goals to be pursued by a people whose thoughts and beliefs are freed from governmental interference. But they are not values to be implemented by judicial intrusion into otherwise legitimate state activities. (411 U.S. 1, 36)

In citing social science evidence—for example, the relationship of schooling to voting behavior—the minority exposed itself to a kind of double jeopardy. If the data had been demonstrably faulty or ambiguous, the majority could have rejected the contention that education was fundamental on these grounds alone. It had the additional option of disregarding the evidence altogether by declaring it irrelevant to constitutional issues raised by the case. Since the existing data were conclusive and supported the minority position, Justice Powell, as a matter of artifice or conviction, chose to bypass the factual issues. A similar strategy, somewhat altered, was adopted by Justice Marshall when he denied that the uncertain relationship between costs and educational quality was germane to the resolution of the issues in *Rodriguez*.

Inputs or outputs as a measure of quality? Although the majority chose not to confront some kinds of evidence, it nevertheless insisted that even if education were deemed a fundamental right, the appellees must prove that differential per-pupil expenditures differentially affected educational quality and hence denied equal protection. At every stage of the appeal, from preliminary briefs to oral arguments before the Supreme Court, Texas relied on the research of James Coleman and others which questioned the connection between educational resources and educational achievement (Coleman et al., 1966). The data seemed to show that when the individual and social characteristics of pupils were held constant, the educational system had relatively little impact on measures of cognitive achievement. Variations in school factors seldom overcame variations in family background. Since these conclusions were necessarily based on the observation of existing and possibly inadequate schools, these data could

easily have supported an argument for massive spending to improve schools. The more popular implication drawn from them has been that "money does not matter." Attorneys for the state of Texas, needless to say, preferred the latter view. If educational resources have no effect on achievement, then children who attend schools with antiquated plants, high student-teacher ratios, and poorly trained faculties are not thereby denied the benefits of quality education.

The attorneys for Rodriguez responded to this line of reasoning by assailing the evidence itself and claiming that even if the conclusions of the Coleman school were valid, they were irrelevant to the constitutional issues. Thus, for example, the *amicus curiae* brief submitted by the National Educational Association notes that

Texas would pass off the great inequalities in educational funding with the argument, made for the first time on appeal, that money really does not have an effect upon the quality of education that a child receives. This argument not only defies common sense and the judgments of parents, educators, school administrators, and taxpayers in Texas who in the past decade more than doubled the per pupil expenditures on a statewide basis, but also ignores a substantial body of research demonstrating a positive relationship between educational funding and learning, whether the latter is measured solely in terms of cognitive achievement or in terms of other qualities such as creativity and socialization. (p. 13)

The brief for the plaintiffs states the issue even more confidently. The link between cost and quality of education is "clear, unequivocal and uncontroverted" and refers to the "cynical" allegation by the state of Texas that "money makes no difference in the education of those who have suffered from discrimination, but large sums are required to educate those who are privileged to live in wealthy districts" (p. 23).

The major argument of Justice Marshall's minority opinion is not, however, addressed to the validity of competing versions of the nature of the cost-quality equation. Instead, he makes the legal point that "the burden of proving that these disparities do not in fact affect the quality of children's education must fall upon the appellants. Yet appellants made no effort in the district court to demonstrate that educational quality is not affected by variations in funding and in resulting inputs. And, in this court, they have argued no more than that the relationship is ambiguous" (411 U.S. 1, 86). Moreover, Marshall maintained that, in any event, the cost-quality relationship should have no bearing on the *Rodriguez* decision. The appellees have no legal obligation to demonstrate that educational outcomes are a function of expenditures, but only that resources available to children are unequally distributed. Since the property-poor districts were disadvantaged with respect to physical facilities, teacher salaries, and educational resources, Marshall concludes that this situation, directly attributable to the Texas financing scheme, raises a grave constitutional question of state-created discrimination.

As with other issues before the court, evidentiary disputes on the cost-quality question were essentially peripheral to the central thread of the argument. Justice Powell could have accepted the allegation that cost determined quality—without retreating from his view that the Texas school-financing system was constitutional. Justice Marshall could and did conclude that disparities in opportunities rather than differences in outcomes were sufficient to establish a violation of the Fourteenth Amendment. The evidence was thus employed by both majority and minority to add force and flourish to arguments which, in their opinion, could stand alone even if the social science findings had been reversed.

The definition of equal education. The majority identified educational outcomes and the minority educational investments as the appropriate measure of quality of schooling. Both sides were then obliged to specify the meaning of equality with respect to their chosen criterion. In the words of Justice Powell:

Whatever merit appellees' argument might have if a State's financing system occasioned an absolute denial of educational opportunities to any of its children, that argument provides no basis for finding an interference with fundamental rights where only relative differences in spending levels are involved and where—as is true in the present case—no charge fairly could be made that the system fails to provide each child with an opportunity to acquire the basic minimal skills necessary for the enjoyment of the rights of speech and of full participation in the political process. (411 U.S. 1, 37)

In other words, a state is not obliged to inquire who has mastered calculus so long as all have learned arithmetic. As might be expected, Justice Marshall questioned what judicially manageable standards might be employed to define "how much education is 'enough' to excuse constitutional discrimination" (411 U.S. 1, 89), especially since the majority had professed doubt of any clear relationship between cost and quality. A more generous standard, the adoption of school-inputs that varied according to the individual needs of each pupil, had already been rejected as unfeasible by a previous decision of the Court (*McInnis v. Shapiro*, 293 F. Supp. 327 [WD Ill. 1968] affd. per curiam sub nom. *McInnis v. Ogilvie* 394 U.S. 322 [1969]). Under these circumstances, Marshall concluded that

appellees do not now seek the best education Texas might provide. They do seek, however, an end to state discrimination resulting from the unequal distribution of taxable district property wealth that directly impairs the ability of some districts to provide the same educational opportunity that other districts can provide with the same or even substantially less tax effort. . . . As this court held in *Brown v. Board of Education* . . . the opportunity of education "where the state has undertaken to provide it, is a right which must be made available to all on equal terms." (411 U.S. 1, 115–116)

Since all Texas schools presumably provide basic skills, the majority position had the effect of defining equality in expenditures as a sum equal to the spending level of the poorest districts. By contrast, the minority

defined a standard of equality that required comparable expenditures for all children and all districts. Here, as elsewhere, the debate could have been conducted philosophically, without recourse to facts about the existing situation. The court could have judged this issue purely on its judicial merits even if it had no access to evidence.

Rational or compelling state interests: Social consequences of an educational financing system based on the local property tax. The ruling of the majority on other aspects of the *Rodriguez* case in effect relieved it of any real necessity to show that Texas' educational policy satisfied a compelling state interest. In the absence of a suspect category or a violation of a fundamental right, Justice Powell held

that the constitutional standard under the Equal Protection Clause is whether the challenged state action rationally furthers a legitimate state purpose or interest. We hold that Texas' plan abundantly satisfies this standard. (411 U.S. 1, 55)

A state interest that particularly concerned the majority was the maintenance of local control over educational policy. The majority position was that he who pays the piper calls the tune and that the doctrine of fiscal neutrality would thus severely restrict such traditional local prerogatives as hiring teachers, establishing curriculum, and adopting budgets that reflect the community's wishes.

The Court seems to have been impressed by the line of reasoning perhaps best stated in an *amicus curiae* brief submitted by the superintendent of schools, Los Angeles County.

A further interest of society and an individual interest of parents which would be affected by affirmance of the rule adopted by the court below has recently been recognized and reaffirmed by this court. This is the interest of the parent in being able to control his child's education and upbringing. . . . The constitutional rules urged by plaintiffs and adopted by the lower court would effectively limit the opportunities within the public school system of a parent who desired to pay more for his child's education or to have an effective voice in the determination of the amount of funds to be expended on educational programs within his local school district. (*San Antonio v. Rodriguez*, Brief of *Amicus Curiae*, Richard M. Clowes, Superintendent of Schools of the County of Los Angeles et al., pp. 35-36)

In meeting these arguments the minority insisted that the state must show a "compelling," not merely a "rational," interest in maintaining a system that was based on wealth discriminations and violated the fundamental rights of children. In doing so, the dissenters did not flatly deny that local control of schools might well represent a compelling interest but chose instead to rest their case on factual grounds. They contended that local control of educational policy would not be substantially affected by increased state assumption of financial support for schools, especially since the appellees did not propose outright abolition of the local property tax. In Justice Marshall's opinion, even fully "centralized educational

financing . . . would leave in local hands the entire gamut of local educational policymaking . . . teachers, curriculum, school sites, the whole process of allocating resources among alternative educational objectives" (411 U.S. 1, 131, n. 98). Indeed, Justice White's dissenting opinion held that the Texas system was culpable precisely because interdistrict disparities in taxable income severely restricted local options.

The difficulty with the Texas system, however, is that it provides a meaningful option to Alamo Heights and like school districts but almost none to Edgewood and those other districts with a low per-pupil real estate tax base. In these latter districts, no matter how desirous parents are of supporting their schools with greater revenues, it is impossible to do so through the use of the real estate property tax. In these districts the Texas system utterly fails to extend a realistic choice to parents, because the property tax, which is the only revenue-raising mechanism extended to school districts, is practically and legally unavailable. That this is the situation may be readily demonstrated. (411 U.S. 1, 64, 65)

In retrospect it appears that the minority substantially weakened its case by failing to challenge the sanctity of local control. It might have chosen instead to assert that in public education, fiscal neutrality and equal educational opportunity were paramount state interests that superseded all others. Instead, the dissenting opinion in effect accepted the rational interest test and was content to claim only that the factual evidence would not sustain even this modest standard. However, since competing views of the relationship between state financing and local control were argued on the basis of plausible conjecture, the dissenting opinion could hardly have been overwhelmingly persuasive. Detailed, incontrovertible empirical accounts of the Hawaiian system might have demonstrated that centralized financing and local educational options do not inherently conflict, but it is not difficult to imagine the subsequent jousting about the generalizability of the case, the adequacy of the sample, the nature of the research design, etc. Empirical evidence at that time, even under the best foreseeable circumstances, could have been contested by the majority and would have provided scant comfort to the minority.⁸

Are there alternate feasible methods of financing schools? The majority opinion of the Court was not required to suggest an alternative system of school finance, since in its view the Texas system already met the requirements of the Fourteenth Amendment. Justice Powell's position on this matter was thus in the nature of *obiter dicta* through which the more

⁸ In a challenge to Connecticut's school-finance law argued in 1974, the trial court heard oral expert testimony on school-finance reform laws enacted in other states during the previous year which were variants of power equalization and thus permitted local choice over tax rates. In striking down the Connecticut system which did not provide equalization of differences in local wealth, the Court concluded: "There is, however, no reason why local control needs to be diminished in any degree merely because some system other than the present system is adopted. Indeed, there was convincing evidence that other systems have been adopted without any loss of local control" (*Horton v. Meskill*, Superior Court, County of Hartford, No. 18 52 83, December 26, 1974, p. 12; *aff'd* on appeal, April 19, 1977).

"conservative" justices, most of them Nixon appointees, affirmed their commitment to a less activist stance than had characterized the Warren Court. In concluding his opinion, Justice Powell wrote:

The need is apparent for reform in tax systems which may well have relied too long and too heavily on the local property tax. And certainly innovative thinking as to public education, its methods, and its funding is necessary to assure both a higher level of quality and greater uniformity of opportunity. These matters merit the continued attention of the scholars who already have contributed much by their challenges. But the ultimate solutions must come from the lawmakers and from the democratic pressures of those who elect them. (411 U.S. 1, 58, 59)

Justice Marshall, whose career on the court can be described as interventionist, responded in characteristic fashion. "I, for one, am unsatisfied with the hope of an ultimate 'political' solution sometime in the indefinite future while, in the meantime, countless children unjustifiably receive inferior educations that 'may affect their hearts and minds in a way unlikely ever to be undone'" (411 U.S. 1, 71-72). At the same time, if tax reform were to cause chaos in the school-finance systems of Texas and, by judicial extension, other states, a compelling state interest in the current system could be alleged. For this reason Marshall suggested a number of plans—centralized education financing, district power equalization, district wealth reapportionment, and others—which "indicate the breadth of choice which remains to the state if the present inter-district disparities were eliminated" (411 U.S. 1, 131). Unfortunately, the literature of scholarly reform did not, at that time, contain conclusive evidence that any of these proffered solutions were feasible. Indeed, as the appellants noted before the district court, the two most celebrated names in the reform movement are highly skeptical of each other's proposals:

Professor Wise and those aligned with him think that "district power equalizing" would be unconstitutional while Professor Coons and his associates are almost equally critical of the solution pushed by Professor Wise. (*San Antonio v. Rodriguez*, Brief for the appellants, p. 16)

And in similar spirit, Justice Powell was careful to footnote his concluding description of the high costs and the problems for urban areas that might occur under potential school-finance systems with articles written by the principal fiscal expert for the plaintiffs (411 U.S. 1, 56, 57, 58, n. 111, 113, 114).

It is apparent from the preceding discussion of how the Justices of the Supreme Court resolved the legal and factual issues raised by the *Rodriguez* case that evidence of any kind was frequently irrelevant to the argument; that where it was relevant, it was selectively used; and that once introduced, social science proved an uncertain ally in debate. It would be easy but incorrect to deduce from this that social research had no appreciable effect upon the Texas litigation and negligible impact in other

school-finance cases or legislative developments. The remainder of this chapter may correct this hasty impression.

4. SOCIAL SCIENCE, THE LAW, AND EDUCATIONAL REFORM: SOME OBSERVATIONS ON JUDICIAL REASONING

The law may be viewed as a set of codified norms that limit the freedom of judges to render decisions resting on their own philosophical preferences or their private conceptions of desirable outcomes. This is the root meaning of the familiar observation that "we are a government of laws, not men." The reactions of peers, the legal profession, public opinion, and, even more important, the judge's anticipation of the social consequences of his ruling also temper personal predilections. The two chief constraints on the judiciary are a body of law consisting of legal principles and the rules of evidence that govern the establishment of legal facts.

4.1 *The Sources of Authority*

The legal principles governing the *Rodriguez* case were established as elsewhere by reasoning from precedents established in other cases. As Edward Levi has noted, this process consists essentially of observing the similarities between two cases, stating the rule of law governing the first, and applying this rule to the second (Rosen, 1972, p. 11). The factual base underlying the *Rodriguez* decision rested on three sources of evidence: common knowledge, judicial authority, and social research. Common knowledge consists of "self-evident" assertions requiring no additional authority because they have presumably been ratified by ordinary experience and the collective wisdom of mankind. Judicial authority consists of facts explicitly stated or derived from previous judicial opinions or from legal scholars, without reference to other forms of proof. Social research consists of data generated by professional social scientists, mainly academics, whose claim to a hearing rests on their presumed competence in one or more disciplines and their adherence to the canons of scientific method.

Some illustrations drawn from *Rodriguez* may clarify these three kinds of evidence.

Common Knowledge:

Likewise it is difficult to believe that if the children of Texas had a free choice, they would choose to be educated in districts with fewer resources, and hence with more antiquated plants, less experienced teachers, and a less diversified curriculum. In fact, if financing variations are so insignificant to educational quality, it is difficult to understand why a number of our country's wealthiest school districts, who have no legal obligation to argue in support of the constitutionality of the Texas legislation, have nevertheless zealously pursued its cause before this Court. (411 U.S. 1, 85)

Judicial Authority:

The merit of local control was recognized last Term in both the majority and dissenting opinions in *Wright v. Council of the City of Emporia*, 407 U.S. 451 (1972). Mr. Justice Stewart stated there that "[d]irect control over decisions vitally affecting the education of one's children is a need that is strongly felt in our society." *Id.*, at 469. The Chief Justice, in his dissent, agreed that "[l]ocal control is not only vital to continued public support of the schools, but it is of overriding importance from an educational standpoint as well." *Id.*, at 478. (411 U.S. 1, 49)

Social Science:

This is not the place to weigh the arguments for and against "district power equalizing," beyond noting that commentators are in disagreement as to whether it is feasible, how it would work, and indeed whether it would violate the equal protection theory underlying appellees' case. President's Commission on School Finance, *Schools, People & Money* 32-33 (1972); Bateman & Brown. Some reflections on *Serrano v. Priest*, 49 J. Urban L. 701-708 (1972); Brest, Book Review, 23 Stan. L. Rev. 591, 594-596 (1971); Goldstein, *supra*, n. 38, at 542-543; Wise, School Finance Equalization Lawsuits: A Model Legislative Response, 2 Yale Rev. of L. & Soc. Action 123, 125 (1971); Silard & White, Intrastate Inequalities in Public Education: The Case for Judicial Relief Under the Equal Protection Clause, 1970 Wisc. L. Rev. 7, 20-30. (411 U.S. 1, 42, n. 85)

The court in the *Rodriguez* case was notably more inclined to rely on judicial authority than on any other source of support and it did not hesitate to find truth in propositions that rely mainly on conventional wisdom. This is hardly astonishing; judges must rule and they cannot always afford the grace of epistemological modesty. No calendar would ever be cleared if in each proceeding the courts were to transform themselves into research institutes. The social scientist who wishes to influence the judicial process is thus obliged to inquire, first, under what circumstances any fact, regardless of the authority on which it rests, will be a controlling factor in resolving legal issues and, second, to what extent evidence from the social sciences is demonstrably superior to other sorts of knowledge.

4.2 The Criterion of Relevance

As we suggested, the Supreme Court's majority and minority opinions both exhibited remarkable agility in accepting those putative facts they required for an argument and disregarding as irrelevant other data that might derail an otherwise persuasive train of thought. If Justice Powell, for example, did not find it germane that school expenditures vary systematically according to district property wealth or if Justice Marshall found the cost-quality equation beside the point for equal-protection analysis, even the most exquisitely wrought research could not have affected their judgments about these issues in *Rodriguez*. It is true, of course, that some facts are capable of reducing or virtually eliminating judicial

indeterminacy. If the correlation between district school expenditures and the economic characteristics of San Antonio's population had been perfect, or at least extraordinarily high, the majority as well as the minority might well have affirmed that the resulting discrimination was based on the "suspect" category of wealth. But social science findings are seldom wrapped in airtight packages and it is almost always possible to use evidence selectively to deny the relevance even of facts that are beyond dispute.

The status of evidence vis-à-vis legal principles is especially low when a litigant wishes to demonstrate that no "rational" or "compelling" state purpose is served by the action of a legislature. When in 1905 Louis D. Brandeis first made extensive use of social science in what has come to be known as the "Brandeis brief," he was defending Oregon's right to regulate the working hours of laundresses. Although his motives were indubitably "liberal," he was, in effect, supporting "states' rights" against the intrusion of excessively interventionist courts (Rosen, 1972, p. 72 and *passim*). In submitting a massive array of empirical evidence presumably demonstrating physiological differences between the sexes and the greater susceptibility of women to fatigue, Brandeis did not contend that his data were decisive but only that the law was based on plausible, if inconclusive, assumptions. Paradoxically, Brandeis thereby forged a legal instrument which in most contexts favors the status quo more than social change. Once a state such as Texas is able to show that its school-finance system rests on some rational grounds, all counterarguments that assert the equal or superior rationality of competing legislative strategies are, by the terms of the Brandeis brief, irrelevant.

4.3 *The Limitations of Traditional Social Science in the Judicial Process*

The impact of social science on educational law, then, is weakened by the fact that one of its principal exports, empirical evidence, simply may be judged not germane to the case. But even when a class of facts is admittedly relevant, the claims of social science may be discounted by the courts because of the state of the art. Ideally, a social-scientific proposition introduced into a legal proceeding should be clear, certain, stable, and framed at the appropriate level of abstraction. These desiderata are seldom fully realized. Practitioners of social science have, after all, mastered a demanding craft, sometimes use an arcane technical vocabulary, and employ sophisticated quantitative methods. It is difficult to imagine that these modes of thought are congenial to most courts or, indeed, that jurists always comprehend them.

In "Social Science and the Courts: The Detroit Case," Eleanor P. Wolf (1976) illustrates how differences in training and outlook can diminish the value of social science testimony:

There were lengthy statistical presentations on education-resource allocations to predominantly black and white schools presented along with school achievement levels, but no explanation of how to discover what causal relationship, if any, exists between variations in school inputs and variations in learning. Indeed, during the remedy hearings (after the trial ended) the judge revealed that he had thought all along that the child's socio-economic status was "part of input"! What had he made of all those discussions about resources and social class? It seems clear that much of the significance of this key causal controversy somehow escaped him, yet his intelligence and conscientious interest are clear from the record. (pp. 112-113)

An equally revealing instance of the failure of communication appears in the testimony of Roger T. Lennon as an expert witness on psychological testing in the District of Columbia schools case of *Hobson et al. v. Hansen et al.*:

Q Isn't there another reason, Dr. Lennon? Didn't you discover that the Otis Quick-Scoring Test is a fairly unreliable test?

A No, sir. I think that you are referring to my report that we have discovered that it yields I.Q.s that we judge to be six or seven points too high, higher than they would be if we normed it on today's population. This again does not run to the reliability, as we use the term, of the results.

Q Do you mean to tell me that a test is showing an I.Q. that is seven points too high, that you would classify that as a "reliable" test?

A In the nomenclature of the discipline, Mr. Kunstler, that is not what we connote by the term "reliable." *Reliable* refers precisely to the stability of the result, not to its closeness to reality, if you will, or its accuracy.

Q Dr. Lennon, if I understand what you are saying, "reliability" in the nomenclature of the profession has no relationship to accuracy?

A To the accuracy of the norms. That is exactly correct.

Q And that would be the same if the I.Q.s were shown as seven points too low?

A Or 47 points too high or too low.

Q I see.

A I may say, Mr. Kunstler, that our efforts to have this concept of reliability as we use it in our jargon, if you will, generally understood, are a good deal less than successful. Your usage corresponds much more to the man in the street's or even to that of _____.

Q Now, Dr. Lennon—let me withdraw that. (Lennon, n.d., p. 112)

Expressing his frustration after having listened to expert witnesses for 19 months in this case, Judge Skelly Wright, a federal judge widely known for his thoughtful, precedent-setting opinions, complained that he had had to fall back on "simple arithmetic and morality" in making his decision.

The unfortunate if inevitable tendency (in this debate) . . . is to lose sight of the disadvantaged young students on whose behalf the suit was brought in an overgrown garden of numbers and charts and jargon like "standard deviation of the variable, statistically significant Pearson product moment correlations." The reports by the experts . . . are less helpful than they might have been for

the simple reason that they do not begin from a common data base, disagree over crucial statistical assumptions, and reach different conclusions. . . . The lawyers in this case had a basic responsibility which they have not completely met to put the hard core statistical demonstrations into language which serious and concerned laymen could, with effort, understand. (*Hobson v. Hansen*, 327 F. Supp. [D.D.C., 1971] at 859)

The courts might trouble to learn the language and techniques of social science if research yielded definitive findings that were demonstrably valid. Social science is most successful in this respect at the descriptive level and when it employs simple variables with straightforward meanings. It was comparatively simple to show that in San Antonio an inverse association existed between district school expenditures and local tax effort and that this relationship was especially strong at the extreme ends of the income distribution. This demonstration impressed the district court which ruled in favor of the plaintiffs and strongly influenced at least four members of the Supreme Court. Indeed, without such evidence it is difficult to imagine how the issues raised by *Rodriguez* could ever have come before the courts.

Social science is, however, less able to deal with complex and subtle concepts like "equality" and "quality," where operational definitions are elusive and measurement problems formidable, and it is almost wholly incapable of producing dependable general theories at a high level of abstraction. Little wonder that a dismayed court, accustomed to deriving permanent verities from immutable legal principle and observing the professional disagreement caused by the Coleman Report, withdrew to the familiar comforts of judicial authority and common knowledge. This withdrawal is even more likely to occur when a decision partly rests on the anticipation of future consequences. The office of social theory is to lessen reliance on raw empiricism by generating deductive propositions that permit us to predict events of the same class, including those that have not yet occurred. Since this degree of intellectual economy is far beyond reach today, the courts are free to rely instead on their own projections of what might happen if, for example, current schemes of taxation were revised.

Social science, then, does not and should not have presumptive claim over other sources of fact except when particular propositions actually meet scientific criteria of proof. Opinions of social scientists that are otherwise uncorroborated by well-confirmed theory or research should have the same legal status as the reflections of other thoughtful people. The values of a scholar inevitably intrude on his work, and the academic community which is overwhelmingly "liberal" and "enlightened" collectively embraces beliefs that are arguably as responsive to a shifting *Zeitgeist* as they are to cumulative research findings. Only recently psychiatric opinion held that women found their true destiny in rearing children; currently the psychiatric literature occasionally gives the impres-

sion that offspring are awkward impediments to achieving the fully realized life. Enlightened criminologists who once endorsed rehabilitation and treatment, indeterminate sentences, and equity proceedings for juveniles have now reversed themselves and adopted positions that 20 years ago were held by their "right wing" adversaries.

Similar transformations have occurred in the theory and practice of education. Homogeneous grouping, once conceived as "individuation," is now "tracking," which is thought to damage the educational careers of children from working-class homes. Standardized tests, once regarded as guardians of egalitarian values, are often perceived today as instruments of aristocratic privilege. Academic tenure, once indispensable to the protection of vulnerable teachers, is now often condemned as an obstacle to community control of schools. The courts may be pardoned if they stand aloof from so fickle a science.

Social science findings are especially unreliable as guides for devising effective means to improve the life chances of children from low-income and minority backgrounds. Henry M. Levin (1976), for example, finds only weak or inconclusive evidence to support the existence of various mental capacities and social goals ("cognitive skills," "non-cognitive work characteristics," "sorting and selection," and "reduction of racial and class frictions") that prompted the development of compensatory education and social action programs. Yet social scientists as legal advocates introduce such evidence as if it could resolve disputes over what are essentially moral problems. According to Levin, referring here to post-Rodriguez litigation,

Presentation of evidence on the relationship between educational expenditures and cognitive achievement implicitly narrows the context within which the effects of unequal expenditure patterns will be considered. While the two sides to this debate disagree on the effect of school resources, both have accepted the view that standardized achievement scores are the appropriate focus for exploring educational outcomes. Thus, much of the legal debate surrounding the challenge to present methods of financing education does not address the basic unfairness reflected by state arrangements to spend more on the education of children in rich districts than in poor ones. *Rather, the courts have become immersed in arguments over whether money makes a difference for "poor kids."* It is unfortunate that the issue has now become framed in terms of whether additional expenditures for children in poor school districts will raise their test scores. (pp. 29-30)

There are thus ample grounds for questioning the power and beneficence of social research as an agent of educational reform through the law, and the case of *San Antonio v. Rodriguez* illustrated the dilemma. The current infirmities of all social disciplines, the nature of their ambiguous, often inconvenient findings, their vulnerability to faddism, and their capacity for transmuting moral issues into empirical controversies should give pause even to the most dedicated positivist.

5. ATTORNEYS AND SOCIAL SCIENTISTS: PARTNERS IN ADVOCACY

In October 1971, during the heady period following *Serrano*, a group of lawyers prominent in school-finance litigation met to discuss what might be done to extend their initial victories. This colloquy, which was subsequently reproduced in the *Yale Review of Law and Social Action* (Lawyers' Committee, 1971), is remarkable for its disdain of any pretense that the participants were seeking disembodied truth. Everyone at the meeting had long since concluded that the prevalent mode of financing schools was unjust and inequitable; they were assembled to discuss how to persuade the courts to give legal embodiment to this conviction. The conversation was open, frank, unguarded—and it was clear throughout that the group sought to define the scope and limits of social science as an ally in the adversary process. The proceedings represent perhaps the best available illustration of the use of social research for tendentious purposes. Here we see the converse of the ideal of neutral scholarship.

John Coons: Yes, indeed . . . If I read *Serrano* correctly, it certainly does not depend on poverty of the plaintiffs. Although it says this issue is a matter of proof, when you turn the page, it says that in any event, collective poverty is sufficient. Now collective poverty, if it is adequate, is a much simpler thing to demonstrate. We don't have to show what is very difficult to show under the present manner of gathering statistics—that is, the relationship between personal and collective poverty.

John Silard: What do you mean by collective poverty?

Coons: By personal poverty I mean family wealth. By collective poverty I mean district wealth, and we don't know except in certain specific cases that poor people live in poor districts.

Silard: I assume from your complaint that this is true.

Coons: In *Serrano* the court said, fine, go back and prove it, but by the way if you don't prove it, it doesn't matter and collective poverty is enough.

David Kirp: Do you think all courts are going to be as lenient?

Coons: No, in fact that's why in the model complaint we stress both personal and collective poverty. I think it may be that some judges will depend upon personal poverty, but if you can get the broader result, it is better.

Silard: You don't mean by collective poverty that the average income of people living in a community is low. You mean the community has a poor tax base.

Coons: That's right.

John McDermott: There has to be some identification eventually with the fact that the lack of resources does relate to the lack of adequate accomplishments in the indigent schools or the poor district schools.

We haven't got the data right now to prove that if you spend \$2400 on a kid, he gets a better education, but I do know if you spend \$600 on a kid, you can't compete with somebody in a richer district.

Coons: Why is it important to show that nexus?

McDermott: Because I think you need more teachers. I think you need lower class size. I think you need more adequate facilities. I think you have to overcome the pathologies of the environment.

Coons: How can you prove these?

McDermott: By using criteria other than Stanford-Binet scores or reading scores. I think you have to take a new look.

Kirp: The issue of whether the cost of a child's education relates to the quality of his education is important for one reason and that is, judges believe it's important. This argument could lead to the whole quagmire that attempting to prove educational need took us into in the first place. I would suggest either a hard-line approach—not making the connection at all—or Jack's approach, which muddles the issue to some extent. It leaves enough of it there so that if a court wants to pick something up about equality, it can do that.

Sidney Wolinsky: I think it's fair to say that the *Serrano* decision somewhat muddles the issue because we muddled it on oral argument. You can't prove it to a certainty and I agree with you in that sense. If you put yourself in a position where you have to prove it that way, you're in trouble, although you can decide as a perfectly rational decision, to take on the issue, depending in part on your court, and a lot of other factors. But I think that we all could agree that it should be said tomorrow that you may have difficulty in producing overwhelming evidence that there is a positive correlation between dollars spent and quality of educational experience. That's a factor that you have to consider.

John Doar: Does anyone disagree that this can't be proven?

Coons: I guess I would have to know what you mean by proof. I think that putting on expert witnesses who are sufficiently persuasive on the question will, in the legal sense, produce adequate proof, and the judge or jury can decide even if it is a disputed question. Of course, I would prefer to treat it formalistically. I would say that the state had decided that money does matter, by virtue of setting up a system in which you purchase goods and services with money, and so the state has foreclosed itself from even contesting that. But that's a formalistic argument which may or may not work.

Peter Rousselot: It's not formalistic. It's an administrative law theory. This is their system. They have decided to work this way and they are allocating things unfairly.

Coons: It's better than I thought then, but even if you have to prove it, you can prove it with expert witnesses.

Silard: We could turn it around. Instead of saying more money buys a better education, you could say almost no money cannot buy you an adequate education. A lot of people agree that the latter is true, even if they might not agree that the former is. In other words we can ask, "can you be educated for \$400 a year in California?" It's a magic question.

Frank Michelman: I would like to go back to the tactical question of fudging. Sid, do you mean by fudging anything other than the argument that the state, by virtue of having set up a system, is foreclosed from raising any question about whether expenditure determines quality?

Wolinsky: Correct. But the complaint should be drafted in such a way that it would also clearly allow you to introduce evidence on that question and would allow the court to make the leap if it so desires. (pp. 112-113)

The spirit of this exchange is consistent with Alice Rivlin's proposal for a "forensic social science" that abandons any claim that it is "part of the tradition of balanced, objective social science in which the scholar hides (or claims to hide) his personal biases, and attempts to present all the evidence on both sides of a set of questions so that the reader may judge for himself" (Levin, 1976, p. 32; Rivlin, 1973). Scholars would write briefs that conceded as little as possible to the opposition, who in turn would likewise feel free to dispense their version of the ungarbled word in the service of persuasion rather than the "whole truth."

The notion of a "forensic social science" also appeals to Robert E. Lindquist and Arthur E. Wise (1976), who in a post-*Rodriguez* autopsy suggest several lines of research for future litigation involving equal-protection analysis.

The majority's assertion of adequacy also suggests an additional course, a more extensive examination of the relationship between education and other factors. This examination should center on the effects of education on the ability to compete in the labor market and to acquire the skills necessary for effective citizenship. In particular, research on the effects of a child's education on his ability to exercise his first amendment rights could spearhead an effort to prove that education is a constitutionally protected interest. This research would concentrate upon how education contributes to a child's conception of his constitutional rights, how education contributes to effective utilization of those rights, and how education serves the function of instilling an understanding of the principles and operation of the governmental process. Mr. Justice Marshall's opinion strongly suggests that there is a significant relation between education and these first amendment rights, and that the closeness of that relation calls for strict judicial scrutiny. The potential efficacy of this approach is also confirmed by Justice Powell's suggestion that "some quantum of Education" may be "a constitutionally protected prerequisite for a meaningful exercise of either right." In other words, it may be possible to refute the majority's assertion that the Texas system provides "each child with an opportunity to acquire the basic skills necessary for the enjoyment of the rights of speech and full participation in the political process." Until evidence of this type is introduced, it will prove difficult to convince the majority that differences in educational expenditures produce an injury so grievous as to require a radical alteration of the education financing systems of the nation. (pp. 20-21)

At first glance, any conception of social research that calls upon its practitioners to reveal only what is useful in achieving a preconceived goal and that rejects any fealty to "balance" and "detachment" seems antagonistic to any known code of scholarly ethics. Barrington Moore stated the case well (Wolf, 1976):

If a defense lawyer is suddenly surprised by a piece of evidence as an argument turned up by the prosecution he is never supposed to say: "My, I had

never thought of that! It's a good point and my client must be guilty." On the other hand, that is exactly what an intellectually honest scholar is supposed to do under the circumstances. (p. 114)

"Intellectual honesty," however, derives its meaning from specific contexts; it is a relationship between a scholar's behavior and the standards he professes. Thus, what is inappropriate in a seminar may be acceptable in the courtroom, and the social advocate practicing forensic social science is no more deceitful than any other practitioner of adversary law.

What can be done to clarify in practice these separate and distinct roles a social scientist may occupy—scholar and advocate? Numerous procedural mechanisms, including panels of consultants selected from among lists submitted by professional disciplinary associations, might be established to help justices to evaluate conflicting scholarly evidence presented in the course of litigation. The verdict of a judge who was thus protected might then reflect the more impartial norms of academic social science, but he would have no occasion to rebuke social advocates for speaking partial truths.

6. RECENT DEVELOPMENTS IN SOCIAL ADVOCACY, POLICY RESEARCH, AND PUBLIC POLICY ON SCHOOL FINANCE

The research generated by the social advocate, then, can contribute to the judicial process in the four principal areas identified by Arthur E. Wise (1977, p. 2): fact determination, predicting the consequences of decisions, formulating remedies, and evaluating implementations. Moreover, since social advocacy is energized by specific preferences for particular social policies, it can fashion comprehensive strategies and employ varied tactics in seeking its goals. The influence of social advocates did not begin with *Rodriguez* nor did it cease when Justice Powell's majority opinion was issued.

6.1 *Recent Legal Developments*

While the legal challenge to imperfectly equalizing finance laws failed at the federal level, it has succeeded in at least eight state courts. Perhaps more important, its indirect impact on public awareness and political action has been far more extensive. During the 1970s as many as 20 states revised their school-funding systems to achieve much the same ends as were sought in the Texas case. The legal successes in the 18 months between *Serrano I* and the Supreme Court decision in *Rodriguez* had set in motion a host of state commission and legislative committee studies that brought new laws—in some cases because of court orders, in others because the logic of the precedent-setting cases in other states proved

convincing, and in still others because policymakers wished to act before the courts compelled them to do so.

The wave of reform continued in the immediate post-*Rodriguez* years. Although it has subsided somewhat since 1975, this abatement may be more plausibly attributed to increasingly austere state budgets rather than to the unfavorable decision of the Supreme Court. A verdict sustaining the district court in the *Rodriguez* case would have been an unconstitutional triumph of the doctrine of fiscal neutrality, but judicial disapproval of this concept did not mark its final defeat.

It is tantalizing to reflect that a single defection from the ranks of the Court's majority would have transformed the American system of school finance. How close the Court was to that result was underlined by the conclusion of Justice Stewart, the fifth vote for the majority, in his concurring opinion: "The method of financing public schools in Texas, as in almost every other state, has resulted in a system of public education that can fairly be described as chaotic and unjust. It does not follow, however, and I cannot find, that this system violates the Constitution of the United States" (411 U.S. 1, 59). We are thus reminded that the Supreme Court, not justice, is the final arbiter of the law. At the same time the Court's serious consideration of the complex range of educational, legal, and economic issues raised by the *Rodriguez* case undoubtedly stimulated other jurists, legislators, and the public to question the educational status quo. At the very least, the social advocates whose intellectual capital sustained the *Rodriguez* litigation were instrumental in mobilizing reformist energy that otherwise might have remained dormant.

Less than two weeks after the United States Supreme Court decided *Rodriguez* in 1973, the New Jersey Supreme Court in *Robinson v. Cahill* ruled that the state's system of school finance violated the state constitution. In the winter of 1975, a Connecticut trial court also invalidated its state's funding system because it violated state constitutional guarantees of equal protection and free public education, and in 1977 the Connecticut Supreme Court sustained that decision (*Horton v. Meskill*, — Conn. — 1977). In 1977, the California Supreme Court, in what may have been the final round in that protracted *Serrano* litigation, reaffirmed its earlier constitutional theory under the California Constitution rather than the federal Constitution, and approved the lower court decision that had ordered the state legislature to devise a new approach to school finance that would reduce interdistrict spending disparities based on local wealth to no more than one hundred dollars. Education, in the California Constitution as in those of most other states, is an explicitly stated fundamental right, thus triggering strict scrutiny under the *Rodriguez* reasoning applied to the California equal-protection clause.

It is of more than passing interest that the grounds of the California and Connecticut decisions, on the one hand, and the New Jersey ruling,

on the other, are substantially different. The California and Connecticut rulings were based on the particular state's equal-protection clause, the fundamental nature of education, and fiscal neutrality, that is, the impropriety of making a fundamental right dependent upon the size of the local tax base. The New Jersey court, on the other hand, rejected the "fundamental interest" view of equal protection but found that the state constitutional guarantee of a "thorough and efficient" educational system required the state to provide equal educational opportunity for all children. The reasoning of the *Robinson* opinion appeared to adopt the argument of the plaintiffs in the first unsuccessful school-finance case, *McInnis v. Shapiro* (293 F. Supp. 327 [ND Ill. 1968]), that educational resources should be distributed in proportion to educational needs. While several state courts have rejected school-finance challenges, the fact that all state constitutions have equal protection and education provisions of some kind suggests that court calendars are not soon likely to be free of school-finance litigation.

6.2 *Contemporary School-finance Research*

Social research continues its prominent role in this protracted reform drama. Influential studies have been conducted for state and federal agencies, and the resulting reports, articles, and books have been widely circulated among policymakers and interest groups concerned with school-finance legislation. Leading scholars in the field maintain a busy schedule of speeches, conferences, and workshops sponsored by organizations such as the Education Commission of the States, the National Conference of State Legislatures, the U.S. Office of Education, the National Lawyers' Committee for Civil Rights Under Law, the National Education Finance Project, the National Urban Coalition, and the Ford Foundation, to name only the most prominent.

This research and the scholars who conduct it are drawn from all three traditions discussed earlier—school finance, economics, and law. Increasingly, however, their unifying commitment is to what may loosely be termed policy analysis, while their original training and university bases are more often in political science, public finance, and public and business administration. Five types of studies have contributed directly to new policy for school funding.⁹

⁹ Examples of this research are: Benson, 1972 [California]; Commission to Study, 1975 [Connecticut]; Shalala and Williams, principal consultants; Governor's Citizens' Committee, 1973 [Florida]; Kirst and Garms, principal consultants; Hickrod, Hubbard, & Yang, n.d. [Illinois]; Levin, Muller, Scanlon, & Cohen, 1972 [President's Commission]; National Education Finance Project, 1971 [Johns and Alexander, Project Directors]; New York State Commission, 1973 [New York]; Benson and Guthrie, Project Directors; Goettel, 1975 [Ohio]; Guthrie, Kleindorfer, Levin, & Stout, 1971 [Michigan]; Odden & Vincent, 1976 [Missouri]; Berke, Kirst, & Usdan, 1977 [politics of school-finance reform]; Shalala & Williams, 1976 [politics of school-finance reform].

First, analyses of fiscal needs and resources have demonstrated disparities among communities in taxable capacity, tax burdens, and revenue yields in order to establish a legislative rationale for new laws. Further, analyses of substate regional differences and variations in pupil populations have begun to find remedy by generating special tax treatment for some urban areas or special distributional provisions for high-cost pupils, for example, vocational, handicapped, or disadvantaged students.

Second, policy studies have illustrated the failure of traditional school-finance laws to meet these problems. Aid formulas designed for an era of wealthy central cities and impoverished outlying areas no longer serve the world of the 1970s, which has experienced great shifts in economic base. Contemporary policy analysts have shown that traditional provisions, formally neutral in tone—such as state aid based on proportionately matching local salary levels—disproportionately aid wealthy, frequently suburban districts able to attract more highly qualified teachers. They have also demonstrated that formulas that are nominally equalizing frequently have so many limitations upon their operation that, in practice, the pattern of spending is highly and positively correlated with property valuation per pupil. Such studies have been prepared for plaintiffs in post-*Rodriguez* school-finance suits as well as for state legislatures and commissions.

Third, scholars of policy analysis have developed alternative proposals designed to overcome such problems. Drawing upon and adapting traditional school-finance systems, they have given substance to differing value approaches by providing the technical mechanisms for carrying out divergent concepts of equity. Thus, states seeking to maintain local choice of tax rates can provide "equal yield for equal tax effort" among districts, while states that care less for local influence over tax rates but that define equity as "matching educational resources to the needs of different types of pupils" can be assured of finding legislative formulas for implementing their goals as well.

Fourth, policy analysts have adapted computer-simulation technology to project the likely effects of enacting many potential formulas, thus permitting policymakers to draft legislation with considerably better estimates than in the past of likely tax and distributional effects. In addition, by merging school finance, general municipal funding, and demographic data, policy analysts are able to show the impact of such alternatives on a variety of factors that far surpass the basic legislative concern with "how much will my district get?" and "how much will another approach to state aid cost?" The impact upon income groups, geographic regions, urban/suburban/rural locations, ethnic and racial minorities, and other such variables is presented in computer printouts, summary tables, and, finally, staff reports.

Fifth, while assembling coalitions is the particular province of politicians, policy analysts—with an eye to shaping legislation that will be both

substantively sound and politically acceptable—examine the political systems through which such legislation will be enacted and enforced. As experience in different states accumulates, scholars can draw on comparative case studies to note the advantage of the legislative route over the referendum; the importance of connecting study commissions to political dominance; the influence of state history and political culture; the role of extraordinary events such as legal cases; the preeminence of political—not educational—leadership and interest groups on the school-finance issue; the formation of coalitions through compromise or support of related pieces of legislation; and, finally, the impact of the economy on resources available for education.

Thus, for the scholars currently working to draft school-finance reform laws, the prime stocks-in-trade are an ability to focus social science on school-finance policy, the use of computer-based quantitative methods, and sensitivity to political processes.

6.3 *Contemporary Legislation to Reform School Finance*

The recent laws that have been influenced by this research have addressed questions of equity in the raising of revenues and equality of opportunity in their distribution. On issues of both taxation and distribution, the recent state laws reflect the same competing approaches that characterized traditional school-finance formulas and the litigation we have examined. On the tax side, the competing concepts are whether to provide equal yield for equal property tax effort among districts, with districts freely choosing their level of effort, or whether to mandate uniform property tax rates for education in all districts or to use state-wide revenue sources, with little (if any) local choice of taxing or spending levels. On the expenditure side, the laws attempt either to provide nearly uniform per-pupil expenditure in every district or to vary expenditure levels with presumed differences in needs for services based on differences in pupil and district characteristics or (in the case of states with equal-yield-for-equal-effort formulas) to ensure expenditure levels that reflect fiscally neutral local preference.

Even federal aid to education has been affected by recent reforms. While currently no major funding program is under consideration to help states address basic tax and distribution inequities, several federal-aid formulas have been adjusted toward congruence with school-finance reform principles and federal planning money for the development of new state school-finance legislation has also become available.

While school-funding patterns in most states still display the old patterns of taxation and distribution that have stimulated research and policy reform since Elwood Cubberley, significant change has been made in somewhere between ten and twenty states—depending on the definition of “significant”—in the 1970s. There is perceptible progress toward greater equity in taxation, greater equality in basic program provision, and

more emphasis on pupil and district school-finance needs. These achievements are due in no small measure to a generation of social advocates using forensic and policy-oriented social science, for whom setbacks such as the *Rodriguez* case have been skirmishes in a protracted campaign.

APPENDIX A

THE EVIDENTIARY AFFIDAVIT SUBMITTED
BY JOEL S. BERKE IN
RODRIGUEZ V. SAN ANTONIO

INTRODUCTION AND SUMMARY

This affidavit has been prepared as a result of a study of school finance in the State of Texas directed by researchers who have spent the last two years in the analysis of the official financial records of the Texas State Education Department and the Governor's Committee on Public School Education.¹ On the basis of that analysis we conclude that the system of school finance in Texas makes the quality of education a direct function of the wealth of local school districts, providing consistently higher quality schooling in districts with higher property values per pupil and consistently lower quality education in school systems with less local resources available for taxation. Furthermore, our study demonstrates that the poorer districts tax themselves at consistently higher equalized tax rates yet realize far lower tax yields than is true in the richer districts.

Discrimination against the poor is not limited to poverty as measured by property valuation: School districts with lower median family incomes are disadvantaged in regard to the quality of school services. School systems with the highest average incomes have the highest school expenditures; districts with the lowest average incomes have the lowest school expenditures.

Racial discrimination is also readily apparent in Texas educational finance. There is a consistent pattern of higher quality education in districts with higher proportions of Anglos and lower quality education in districts with lower proportions of Anglos. In short, the more Negroes and Mexican-Americans in the school population of a district, the lower its revenues for education.

The effects of these financial patterns have direct impact on the quality of education offered to Texas school children. Both the average salary levels, which are reflective of the quality of teachers, and the number of professional personnel per pupil vary directly and markedly with the revenues available per pupil. Thus, in general, the higher the revenue for the schools, the higher the quality of education in the district.

Our study also shows that responsibility for bringing about this invidious pattern of educational resource allocation rests squarely with the state for two reasons. First, the creation by the state of local districts with unequal taxable resources plays the primary part in establishing a system biased in favor of richer districts. Second, by adopting a state aid system that is insufficiently equalizing to offset the disparities among districts, discrimination against the poor in the provision of school service is assured. In the eyes of the experts who participated in our study, a system more effectively designed to assure inequality of education opportunity would be difficult to design. Yet as the data which follow will indicate, just

¹ For a description of the procedures and methodology used in this study, please see Appendix [omitted].

such a system currently exists in the State of Texas as a whole and in Bexar County in particular.

DISTRICT WEALTH, INCOME, MINORITY STATUS, AND SCHOOL REVENUES

The basic measure of district wealth employed in this study is the value of equalized taxable property per pupil in average daily attendance. While from an economic standpoint there are questions as to whether property value is itself an appropriate criterion of a community's ability to raise revenue, we employ that basic measure because property taxation provides virtually all locally raised revenues for education. In addition, as Table I indicates, there is a rough but important correlation between relative levels of property value and relative levels of average income in school districts: Districts high in property value per pupil are high in average income, those that are moderate in property value are moderate in income level, and those low in property valuation are also low in income.

District Wealth and School Revenue

For purposes of analysis, we have grouped the 110 districts into categories of descending amounts of equalized property value per pupil or, in other words, descending order of community wealth. If we compare the average state and local revenues per pupil as shown on Table I with those district wealth categories, a clear and consistent relationship emerges: The richer the district, the higher the revenues available for each pupil's education. Graph I shows how very direct and

TABLE I
THE RELATIONSHIP BETWEEN DISTRICT WEALTH,
INCOME, MINORITY STATUS, AND SCHOOL REVENUES
TEXAS SCHOOL DISTRICTS
CATEGORIZED BY EQUALIZED PROPERTY VALUES,
MEDIAN FAMILY INCOME, AND STATE-LOCAL REVENUE

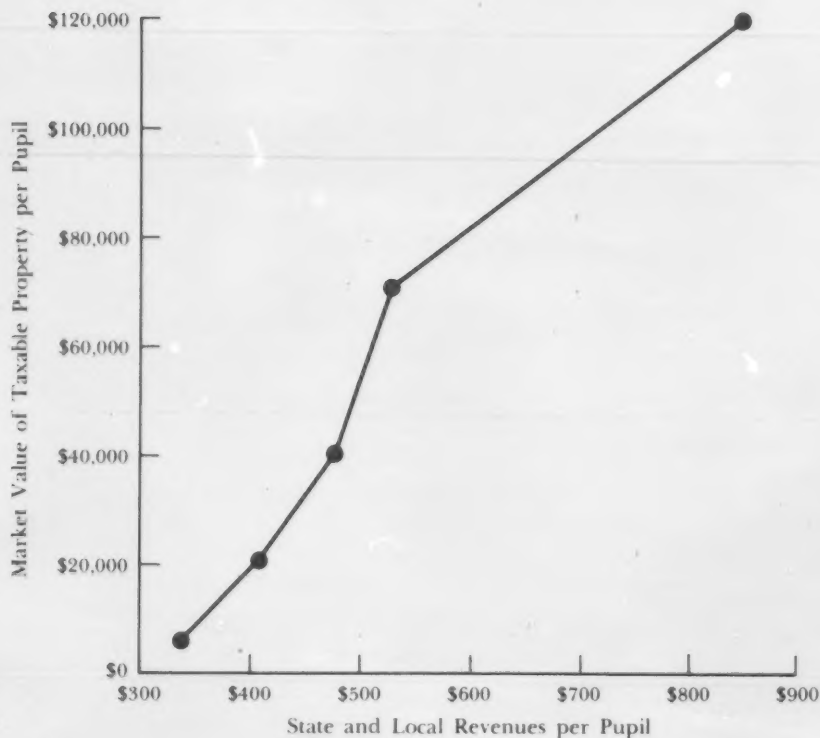
MARKET VALUE OF TAXABLE PROPERTY PER PUPIL	MEDIAN FAMILY INCOME FROM 1960 CENSUS	PER CENT MINORITY PUPILS	STATE-LOCAL REVENUES PER PUPIL
Above \$100,000 (10 Districts)	\$5,900	8%	\$815
\$100,000-\$50,000 (26 Districts)	\$4,425	32%	\$544
\$50,000-\$30,000 (30 Districts)	\$4,900	23%	\$483
\$30,000-\$10,000 (40 Districts)	\$5,050	31%	\$462
Below \$10,000 (4 Districts)	\$3,325	79%	\$305

Reproduced at 411 U.S. 14, n. 38
of Justice Powell's opinion.

Note. Footnotes omitted on all tables in this appendix.

GRAPH I

THE RELATIONSHIP BETWEEN DISTRICT WEALTH
AND STATE-LOCAL REVENUES

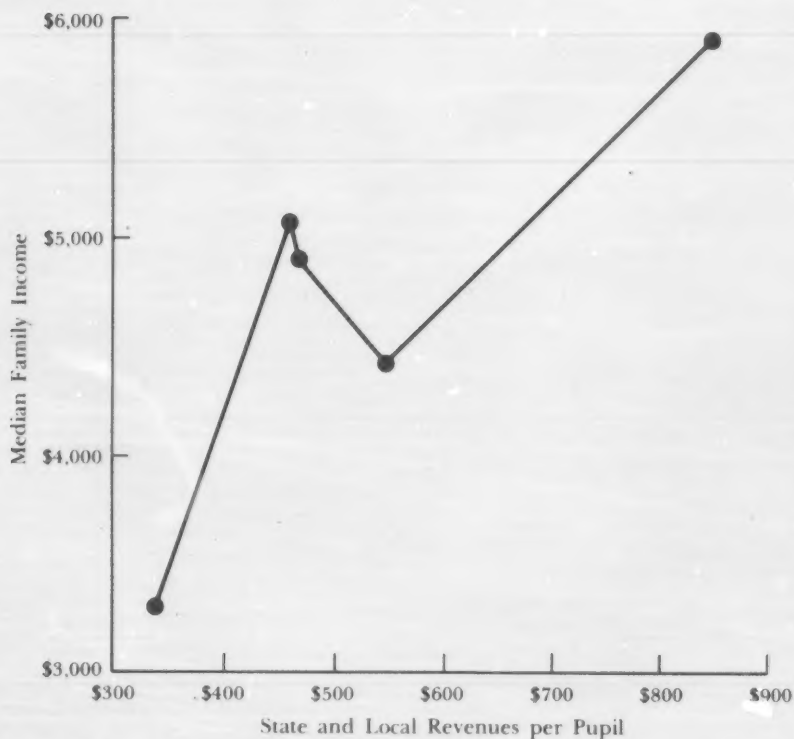


regular the relationship is. Note also the extreme range in both wealth and expenditure categories, from well over \$100,000 to well below \$10,000 in valuation, and from \$815 to \$305 in expenditures—ratios approaching 20:1 in property value and 2:1 in expenditures. The expenditure range in our sample is an average and is therefore a particularly compelling statistic, indicating that a regular pattern exists in which children in poor communities consistently receive an education two and one-half times less costly than children who live in wealthier places. (For the state as a whole, the ratio of the highest expenditure district to the lowest is \$5,334 to \$264, or 20:1.)

Median Family Income and School Revenue

The relationship between income and revenues is also an important one. It shows that it is not simply governmental or jurisdictional poverty that is inversely correlated with school quality, but personal poverty as well. Table I indicates that districts in the highest category of income (\$5,900) receive the most costly education and those in the lowest income category (\$3,325) receive the least expensive

GRAPH II
THE RELATIONSHIP BETWEEN MEDIAN FAMILY INCOME
AND STATE-LOCAL REVENUES



school services. While the relationships near the average are somewhat mixed, they do not work against the prevailing pattern because the range (only \$600 in income) is too small to be meaningful. In short, as Graph II demonstrates, the directness of the association between income and school quality is clear.

Minority Status and School Revenue

The correlation between the proportion of non-Anglos in the schools and the quality of school services is precisely the reverse of the relationship between income and school services. That is, the lower the proportion of non-Anglos, the higher the school expenditures; the higher the proportion of minority group enrollment, the lower the amount of resources devoted to education. Again, as in the income relationship, Table I and Graph III show some small inconsistency in the three middle categories, but the direction of the numbers and of the line on the graph tells an eloquent and vivid story of denial of equal educational opportunity.

A similar relationship was found between the percentage of Mexican-American pupils and school expenditures in a study conducted by the U.S. Commission on Civil Rights (Chart I). Their examination of Texas school systems with at least 10% Mexican-American enrollment showed that, with a minor anomaly in the category of districts with the smallest proportion of Mexican-Americans, school expenditures declined as the proportion of Hispanics increased.

GRAPH III

THE RELATIONSHIP BETWEEN PER CENT MINORITY
AND STATE-LOCAL REVENUES

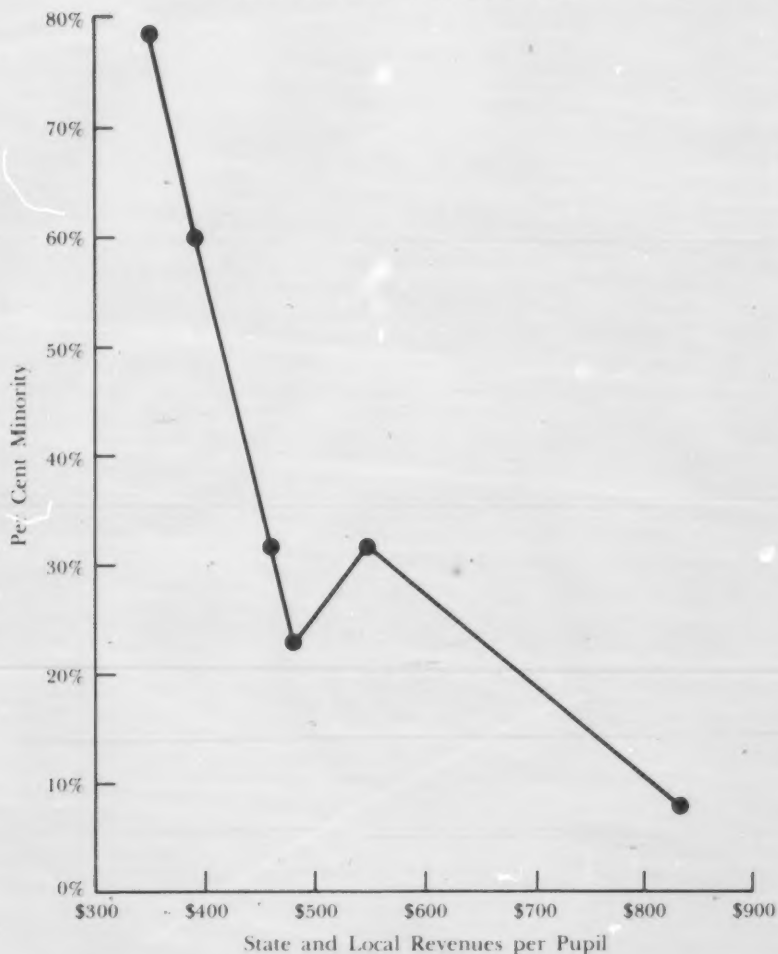


CHART I
EXPENDITURES PER PUPIL IN ADA IN TEXAS
DISTRICTS 10 PER CENT OR MORE MEXICAN-AMERICAN
WITH TOTAL ENROLLMENT 300 PUPILS OR MORE
(Expenditures are from State and local revenue only)

PER CENT MEXICAN-AMERICAN OF DISTRICT ENROLLMENT	DISTRICTS IN SAMPLE		ESTIMATES FOR ALL DISTRICTS	
	Number of Districts	Per Pupil Expenditures	Number of Districts	Per Pupil Expenditures
10-19.9	55	\$457	85	\$444
20-29.9	38	\$484	59	\$477
30-49.9	32	\$444	49	\$444
50-79.9	39	\$377	60	\$382
80-100	23	\$292	30	\$297

DISTRICT WEALTH AND TAX EFFORT

One of the cruel ironies in the current approach to supporting schools in Texas is that the communities that have the least money for their schools are the very districts that tax themselves most heavily to raise school revenues. Using equalized tax rates which permit comparisons between districts, Table II shows an unbroken and consistent inverse relationship between equalized district wealth and effort. The richer a district is, the less severely it need tax itself and—as the third column on the table shows—the more it realizes in locally raised revenues. In short, as tax rates increase, the amount realized decreases.

DISTRICT WEALTH AND HIGHEST TAX EFFORT AND YIELD

Table III takes the relationship between district wealth and tax effort a step further. The table assumes a similar tax effort—the effort made by the district with the highest tax rate in the sample—throughout all districts and examines the resulting differences in yield. The table clearly shows that increased tax effort only magnifies the differences in possible return between wealthier and poorer districts.

Table IV concludes our analysis of the relationship between wealth and effort. It points up not only that increased effort is a futile exercise, as is apparent in Table III, but also that the resulting burden increases at a much greater rate for poorer districts than for richer if they both seek to realize the highest return in our sample.

TABLE II
TEXAS SCHOOL DISTRICTS CATEGORIZED BY
EQUALIZED PROPERTY VALUES,
EQUALIZED TAX RATES, AND YIELD OF RATES

CATEGORIES MARKET VALUE OF TAXABLE PROPERTY PER PUPIL	EQUALIZED TAX RATES ON \$100	YIELD PER PUPIL (Equalized Rate Applied to District Market Value)
Above \$100,000 (10 Districts)	\$.31	\$585
\$100,000-\$50,000 (26 Districts)	\$.38	\$262
\$50,000-\$30,000 (30 Districts)	\$.55	\$213
\$30,000-\$10,000 (40 Districts)	\$.72	\$162
Below \$10,000 (4 Districts)	\$.70	\$60

Reproduced at 411 U.S. 135, appendix II
to dissenting opinion of Justice Marshall.

TABLE III
THE RELATIONSHIP BETWEEN DISTRICT WEALTH
AND HIGHEST TAX EFFORT
TEXAS SCHOOL DISTRICTS CATEGORIZED BY
EQUALIZED PROPERTY VALUE AND YIELD GENERATED IF
HIGHEST TAX RATE IS APPLIED TO ALL DISTRICTS

CATEGORIES MARKET VALUE OF TAXABLE PROPERTY PER PUPIL	HYPOTHETICAL YIELD OF HIGHEST TAX RATE PER PUPIL
Above \$100,000 (10 Districts)	\$2,356
\$100,000-\$50,000 (26 Districts)	\$918
\$50,000-\$30,000 (30 Districts)	\$519
\$30,000-\$10,000 (40 Districts)	\$292
Below \$10,000 (4 Districts)	\$108

TABLE IV
THE RELATIONSHIP BETWEEN DISTRICT WEALTH
AND HIGHEST TAX EFFORT
TEXAS SCHOOL DISTRICTS CATEGORIZED BY
EQUALIZED PROPERTY VALUE AND TAX RATE REQUIRED
TO GENERATE HIGHEST YIELD IN ALL DISTRICTS

CATEGORIES	TAX RATE NEEDED TO EQUAL HIGHEST YIELD
MARKET VALUE OF TAXABLE PROPERTY PER PUPIL	
Above \$100,000 *(10 Districts)	\$.64 per \$100
\$100,000-\$50,000 (26 Districts)	\$1.49 per \$100
\$50,000-\$30,000 (30 Districts)	\$2.58 per \$100
\$30,000-\$10,000 (40 Districts)	\$4.88 per \$100
Below \$10,000 (4 Districts)	\$12.83 per \$100

THE RELATIONSHIP BETWEEN DISTRICT WEALTH AND STATE AND FEDERAL AID

Given these disparities in local revenue-raising ability, how effectively is the Texas equalization aid system working? The answer may be found in Table V. While there appears to be a mild equalizing direction in the state aid system, its equalizing effect fails to operate in favor of the poorest districts. In other words, wealthier districts receive less aid than poorer ones in four of the five wealth categories, but the system provides more aid to the three categories clustered around the average than it does to the poorest class of school districts.

Possibly more important than the system's failure to aid adequately the neediest districts is the mildness of its equalizing tendencies. In a state where state aid provides more money for education than do local revenues, a steeply equalizing financial program could redress the imbalances between districts in local revenue capacity. In Texas that does not happen. State aid provides only \$52 more to the second richest category than to the highest despite a local revenue gap of over \$320. Between the next two categories state aid provides virtually no equalization despite a local revenue gap of more than \$60. As local revenue falls off most steeply, state aid, as already noted, fails to equalize.

State action, then, has created a system of financing the public schools of Texas that is characterized by marked disparities in educational expenditures. Those disparities have arisen because the state has created local districts with unequal sources of revenue and then has adopted a state aid system that fails to overcome those inequities. Federal aid, on the other hand, demonstrates how an equalizing

TABLE V
REVENUES OF TEXAS SCHOOL DISTRICTS
CATEGORIZED BY EQUALIZED PROPERTY VALUES AND SOURCE OF FUNDS

CATEGORIES	LOCAL REVENUES PER PUPIL	STATE REVENUES PER PUPIL	STATE-LOCAL REVENUES PER PUPIL (Columns 1 and 2)	FEDERAL REVENUES PER PUPIL	TOTAL REVENUES PER PUPIL (State-Local- Federal, Columns 1, 2 and 4)
MARKET VALUE OF TAXABLE PROPERTY PER PUPIL					
Above \$100,000 (10 Districts)	\$610	\$205	\$815	\$41	\$856
\$100,000-\$50,000 (26 Districts)	\$287	\$257	\$544	\$66	\$610
\$50,000-\$30,000 (30 Districts)	\$224	\$260	\$484	\$45	\$529
\$30,000-\$10,000 (40 Districts)	\$166	\$295	\$461	\$85	\$546
Below \$10,000 (4 Districts)	\$63	\$243	\$306	\$135	\$441

Reproduced at 411 U.S. 134, appendix I
to dissenting opinion of Justice Marshall.

system can operate. Although the amounts of federal aid are too small to compensate for the disparities that arise from state action, with the exception of the middle category of districts in Table V federal aid flows in greater magnitudes to poorer school systems.

THE RELATIONSHIP OF DISTRICT WEALTH TO EDUCATIONAL QUALITY

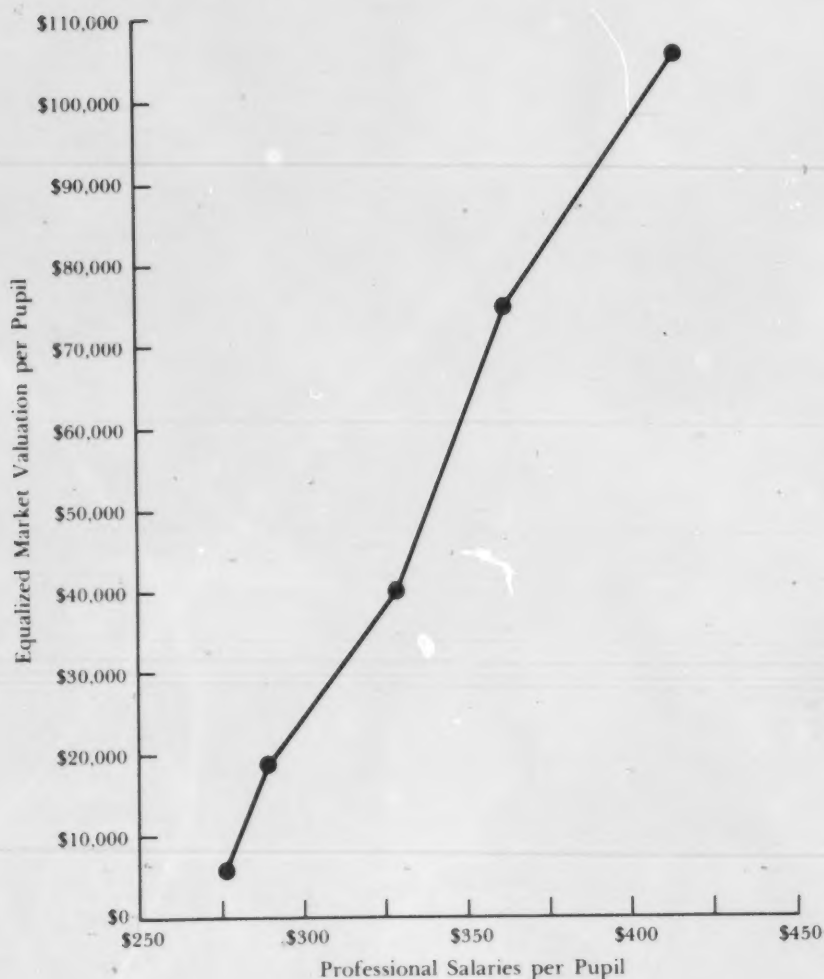
In the preceding discussion we have demonstrated the inequities and disparities that characterize school support in Texas. That analysis has been conducted in terms of revenues, tax rates, equalized property value, and the like. But what is most pressing about the problems we have uncovered is that they have a direct and ascertainable impact on the quality of education afforded youngsters by the State of Texas. While a variety of indicators might be used, we have selected two as examples of the differences in educational quality that exist among school districts.

The first indicator is related to the quality of professional personnel (teachers, guidance counselors, administrators, etc.) in the district. It takes the average cost per pupil for professional personnel, thus subsuming a number of factors including professional training, length of tenure, pupil-teacher ratio, and other less tangible factors. While in individual cases no one would argue that a particular higher salaried professional is more skilled or valuable than one earning slightly less, in aggregate terms and over large numbers of individuals, research suggests that the more a system pays for its staff in comparison with other districts, the better the quality of its personnel. Table VI and Graph IV show the relationship. Without exception, the wealthier the district the higher the professional cost per pupil.

TABLE VI
THE RELATIONSHIP OF DISTRICT WEALTH
TO EDUCATIONAL QUALITY
TEXAS SCHOOL DISTRICTS CATEGORIZED BY
EQUALIZED PROPERTY VALUATION, AND
SELECTED INDICATORS OF SCHOOL QUALITY

CATEGORIES MARKET VALUE OF TAXABLE PROPERTY PER PUPIL	PROFESSIONAL SALARIES PER PUPIL	PROFESSIONAL PERSONNEL PER 100 PUPILS
Above \$100,000 (10 Districts)	\$413.12	\$5.57
\$100,000-\$50,000 (26 Districts)	\$359.72	\$5.17
\$50,000-\$30,000 (30 Districts)	\$327.66	\$4.84
\$30,000-\$10,000 (40 Districts)	\$290.16	\$4.37
Below \$10,000 (4 Districts)	\$276.65	\$4.54

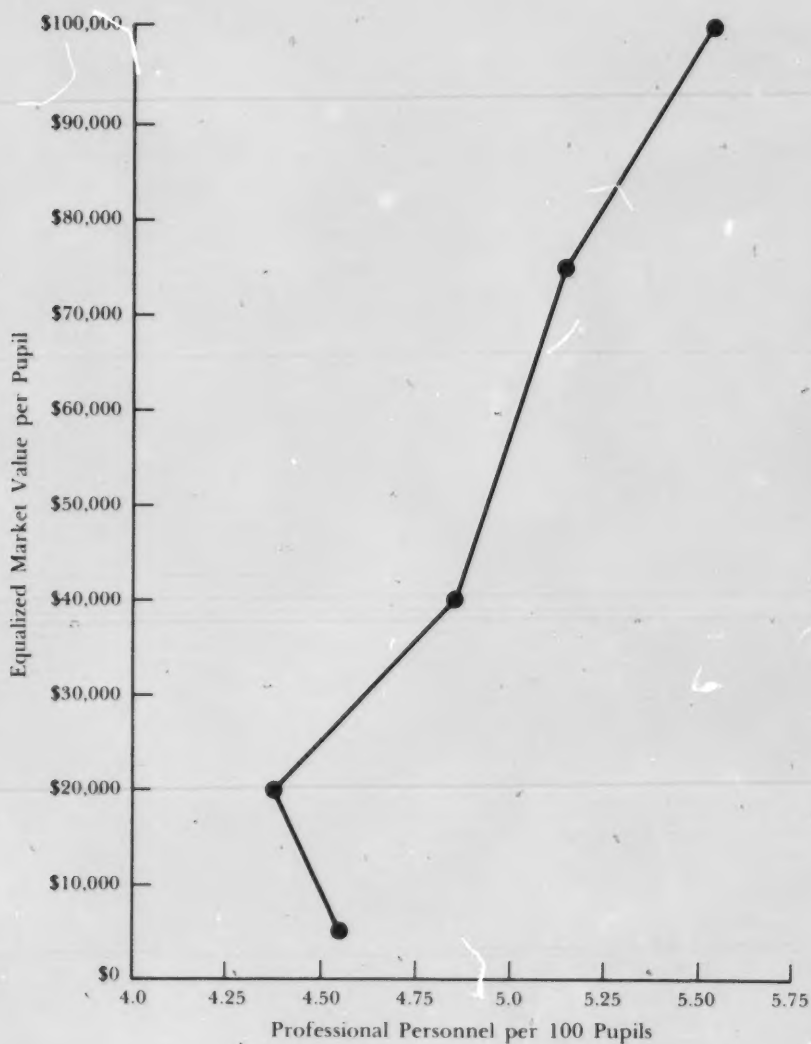
GRAPH IV

RELATIONSHIP BETWEEN PROFESSIONAL SALARIES
PER PUPIL AND EQUALIZED VALUATION PER PUPIL

The second indicator of quality tells virtually the same story. When the ratio of professional personnel to pupils is examined, with the slight exception that the bottom two categories exchange places, the richer the district, the higher the ratio of teachers per 100 pupils. (See Table VI and Graph V.)

In short, higher revenues for education are important because they purchase higher quality education. As we have shown, richer districts not only have the

GRAPH V
THE RELATIONSHIP OF PROFESSIONAL PERSONNEL TO
EQUALIZED MARKET VALUE PER PUPIL



higher revenues but do spend them for more and higher quality teachers, administrators, and guidance counselors.

BEXAR COUNTY

Besides examining Texas school finance through the representative sample of 110 school districts, we have also studied a subsample composed of the six districts in Bexar County which happened to be a part of our larger sample. For a few criteria, we have been able to locate reliable data, and in those instances we have shown relationships for the entire county. In either situation, the six- or the twelve-district Bexar County subsample, patterns do not emerge as clearly as they have for the statewide study. Exceptions and anomalies crop up due to the small number of districts within Bexar County. Yet despite some statistical muddiness, when we conducted an analysis that was parallel to that conducted on the 110-district sample, the conclusions we reached were essentially identical with those we find in the larger study.

DISTRICT WEALTH, INCOME, MINORITY STATUS, AND STATE-LOCAL REVENUE

The relationship between district wealth, income, minority status, and school expenditures in Bexar County holds no surprises to readers of this study. As district wealth as measured by property value declines, so do per-pupil expenditures from state and local revenues. Again, as in the state as a whole, we find a wide range of variation, nearly 10:1 in property values and better than 2:1 in expenditures. While these variations are somewhat smaller than those we found in the state 110-district sample, the fact that variations of that magnitude exist within a single county is remarkable. For the districts on the short end of those discrepancies, the effects are particularly difficult, since they are often in direct competition with their more favored neighbors for personnel and other school resources.

Income variations are both wide and strongly correlated with school expenditures. Table VII shows the familiar Texas pattern: More affluent districts provide consistently more expensive school services. Similarly the relationship between minority status and revenues is as readily apparent in Bexar County as it is in Texas generally. While, as noted above, there are more anomalies in the middle range of wealth in this small sample than we found in the larger, the nature of small, nonrepresentative samples would suggest no other outcome.

DISTRICT WEALTH AND TAX EFFORT

Perhaps the greatest deviation from statewide patterns appears in tax effort. The phenomenon of poorer districts exerting higher tax effort yet realizing lower yields does not emerge as a consistent pattern. Indeed, as Table VIII shows, tax effort appears to be statistically unrelated to district wealth. But this much is clear: There is no suggestion that poorer districts like Edgewood have a lower level of effort than their more fortunate Bexar County neighbors.

DISTRICT WEALTH AND HIGHEST TAX EFFORT AND YIELD

Employing the same analytical technique as we did for the larger sample, however, we find that even if the poorer districts were to expend tax effort equal

TABLE VII

RELATIONSHIP BETWEEN DISTRICT WEALTH,
INCOME, MINORITY STATUS, AND STATE-LOCAL REVENUE
SELECTED BEXAR COUNTY SCHOOL DISTRICTS RANKED BY
MARKET VALUATION, MEDIAN FAMILY INCOME, PROPORTION
OF NONWHITE PUPILS, AND STATE-LOCAL REVENUE

SCHOOL DISTRICTS RANKED FROM HIGH TO LOW BY MARKET VALUE PER PUPIL	MEDIAN FAMILY INCOME FROM 1960 CENSUS	PER CENT MINORITY PUPILS	STATE-LOCAL REVENUES PER PUPIL
ALAMO HEIGHTS (49,478 Market Value per Pupil)	\$8,184	14%	\$558
NORTH EAST (28,202 Market Value Per Pupil)	\$5,900	7%	\$415
SAN ANTONIO (21,944 Market Value Per Pupil)	\$4,691	72%	\$353
NORTH SIDE (20,794 Market Value Per Pupil)	\$4,600	18%	\$362
HARLANDALE (11,345 Market Value Per Pupil)	\$4,436	62%	\$323
EDGEWOOD (5,960 Market Value Per Pupil)	\$3,405	75%	\$248

TABLE VIII
THE RELATIONSHIP OF DISTRICT WEALTH
TO TAX EFFORT AND TAX YIELD
BEXAR COUNTY SCHOOL DISTRICTS RANKED BY
PROPERTY VALUES, EQUALIZED TAX RATES,
AND YIELD OF RATES

DISTRICTS RANKED HIGH TO LOW BY MARKET VALUATION PER PUPIL	EQUALIZED TAX RATES ON \$100	YIELD PER PUPIL (Equalized Rate Applied to District Market Value)
ALAMO HEIGHTS	\$.68	\$343.00
JUDSON	\$.27	\$106.00
EAST CENTRAL	\$.35	\$109.65
NORTH EAST	\$.56	\$173.46
SOMERSET	\$.33	\$87.61
SAN ANTONIO	\$.62	\$136.05
NORTH SIDE	\$.52	\$108.13
SOUTH WEST	\$.27	\$60.19
SOUTH SIDE	\$.39	\$64.61
HARLANDALE	\$.51	\$59.13
SOUTH SAN ANTONIO	\$.61	\$101.73
EDGEWOOD	\$.42	\$26.28

to the highest effort of any district in the county, the resultant differences in yield would consistently favor the wealthy. Thus, in order to realize a yield equal to the highest district in the county, poorer districts would have to expend consistently greater effort than would those with higher property valuation per pupil. This phenomenon is demonstrated in Table IX.

RELATIONSHIP OF DISTRICT WEALTH TO STATE AND FEDERAL AID

When we examine the relationships among different sources of funding in Table X, a crucial effect emerges once again. State aid is unable to offset inequities in relative capacities to finance education that grow out of differences in local taxable resources. The range from \$225 for Alamo Heights, the wealthiest district, to \$250 in Harlandale, the next to the poorest (whose property value per pupil is but one-fifth of that in the wealthiest), is about as mild a pattern of equalization as one could devise. When the fact is that Edgewood with less than one-twelfth the valuation per pupil of Alamo Heights gets only \$3.00 more in state aid, it would appear that there is no meaningful equalization through state resources in Bexar County.

TABLE IX

BEXAR COUNTY, TEXAS, SCHOOL DISTRICTS
RANKED BY EQUALIZED PROPERTY VALUE
AND TAX RATE REQUIRED TO GENERATE
HIGHEST YIELD IN ALL DISTRICTS

DISTRICTS RANKED FROM HIGH TO LOW MARKET VALUATION PER PUPIL	TAX RATE PER \$100 NEEDED TO EQUAL HIGHEST YIELD
ALAMO HEIGHTS	\$.68
JUDSON	\$1.04
EAST CENTRAL	\$1.17
NORTH EAST	\$1.21
SOMERSET	\$1.32
SAN ANTONIO	\$1.56
NORTH SIDE	\$1.65
SOUTH WEST	\$2.10
SOUTH SIDE	\$3.03
HARLANDALE	\$3.20
SOUTH SAN ANTONIO	\$5.77
EDGEWOOD	\$5.76

Reproduced at 411 U.S. 1, 137, appendix IV
to dissenting opinion of Justice Marshall.

TABLE X
THE RELATIONSHIP BETWEEN DISTRICT WEALTH AND SCHOOL REVENUES
REVENUES OF SELECTED BEXAR COUNTY SCHOOL DISTRICTS
RANKED BY PROPERTY VALUES AND SOURCE OF FUNDS

SELECTED DISTRICTS FROM HIGH TO LOW BY MARKET VALUATION PER PUPIL	LOCAL REVENUES PER PUPIL	STATE REVENUES PER PUPIL	STATE-LOCAL REVENUES PER PUPIL	FEDERAL REVENUES PER PUPIL	TOTAL REVENUES PER PUPIL (State-Local- Federal)
ALAMO HEIGHTS	\$333	\$225	\$558	\$36	\$594
NORTH EAST	\$182	\$233	\$415	\$53	\$468
SAN ANTONIO	\$134	\$219	\$353	\$69	\$422
NORTH SIDE	\$114	\$248	\$362	\$81	\$443
HARLANDALE	\$73	\$250	\$323	\$71	\$394
EDGEWOOD	\$26	\$222	\$248	\$108	\$356

TABLE XI
 SELECTED BEXAR COUNTY, TEXAS, SCHOOL DISTRICTS CATEGORIZED BY
 EQUALIZED PROPERTY VALUATION AND SELECTED INDICATORS
 OF EDUCATIONAL QUALITY

SELECTED DISTRICTS FROM HIGH TO LOW BY MARKET VALUATION PER PUPIL	PROFESSIONAL SALARIES PER PUPIL	PER CENT OF TEACHERS WITH		PER CENT OF TOTAL STAFF WITH EMERGENCY PERMITS	STUDENT- COUNSELOR RATIOS	PROFESSIONAL PERSONNEL PER 100 PUPILS
		COLLEGE DEGREES	MASTERS DEGREES			
ALAMO HEIGHTS	\$372	100%	40%	11%	645	4.80
NORTH EAST	\$288	99%	24%	7%	1,516	4.50
SAN ANTONIO	\$251	98%	29%	17%	2,320	4.00
NORTH SIDE	\$258	99%	20%	17%	1,493	4.30
HARLANDALE	\$243	94%	21%	22%	1,800	4.00
EDGEWOOD	\$209	96%	15%	47%	3,098	4.06

Reproduced at 411 U.S. 1, 136, appendix III
 to dissenting opinion of Justice Marshall.

Indeed, what equalization there is among school districts in the area is a function of federal aid, with Edgewood receiving better than three times the allotment of Alamo Heights. Greater amounts of federal aid are virtually perfectly aligned with increasing poverty in the county.

SCHOOL EXPENDITURES AND SCHOOL QUALITY

Table XI displays the striking relationship of district wealth and school quality in Bexar County. Utilizing a wider series of indicators than we did for the larger sample, we find fairly clear patterns of direct correlations between wealth and school quality. In Bexar County as in the state as a whole, the richer school districts appear to be purchasing better quality education for their pupils.

APPENDIX B

EXCERPTS FROM THE OPINIONS IN
SAN ANTONIO V. RODRIGUEZ
OF JUSTICES POWELL AND MARSHALL ON DATA
RELEVANT TO DISTRICT AND PERSONAL POVERTY

*From the Opinion of the Court by Justice Powell
at 411 U.S. 1, 15-16*

Despite these recent increases, substantial interdistrict disparities in school expenditures found by the district court to prevail in San Antonio and in varying degrees throughout the State³⁸ still exist.

³⁸ The District Court relied on the findings presented in an affidavit submitted by Professor Berke of Syracuse. His sampling of 110 Texas school districts demonstrated a direct correlation between the amount of a district's taxable property and its level of per-pupil expenditure. But his study found only a partial correlation between percentage of minorities and expenditures. Although the correlations with respect to family income and race appear only to exist at the extremes, and although the affiant's methodology has been questioned (see Goldstein, *Interdistrict Inequalities in School Financing: A Critical Analysis of Serrano v. Priest and Its Progeny*, 120 U. Pa. L. Rev. 504, 523-525, nn. 67, 71 [1972]), insofar as any of these correlations is relevant to the constitutional thesis presented in this case we may accept its basic thrust. But see *infra*, at 25-27. For a defense of the reliability of the affidavit, see Berke, Carnevale, Morgan & White, *supra* n. 29. [Table I from Affidavit (Appendix A) omitted.]

*From the Opinion of the Court by Justice Powell
at 411 U.S. 1, 26-27*

As suggested above, appellees and the District Court may have embraced a second or third approach, the second of which might be characterized as a theory of relative or comparative discrimination based on family income. Appellees sought to prove that a direct correlation exists between the wealth of families within each district and the expenditures therein for education. That is, along a continuum, the poorer the family the lower the dollar amount of education received by the family's children.

The principal evidence adduced in support of this comparative-discrimination claim is an affidavit submitted by Professor Joel S. Berke of Syracuse University's Educational Finance Policy Institute. The District Court, relying in major part upon this affidavit and apparently accepting the substance of appellees' theory,

noted, first, a positive correlation between the wealth of school districts, measured in terms of assessable property per pupil, and their levels of per-pupil expenditures. Second, the court found a similar correlation between district wealth and the personal wealth of its residents, measured in terms of median family income (337 F. Supp., at 282 n. 3).

If, in fact, these correlations could be sustained, then it might be argued that expenditures on education—equated by appellees to the quality of education—are dependent on personal wealth. Appellees' comparative-discrimination theory would still face serious unanswered questions, including whether a bare positive correlation or some higher degree of correlation⁶¹ is necessary to provide a basis for concluding that the financing system is designed to operate to the peculiar disadvantage of the comparatively poor, and whether a class of this size and diversity could ever claim the special protection accorded "suspect" classes. These questions need not be addressed in this case, however, since appellees' proof fails to support their allegations or the District Court's conclusions.

Professor Berke's affidavit is based on a survey of approximately 10% of the school districts in Texas. His findings, previously set out in the margin, show only that the wealthiest few districts in the sample have the highest median family incomes and spend the most on education, and that the several poorest districts have the lowest family incomes and devote the least amount of money to education. For the remainder of the districts—96 districts composing almost 90% of the sample—the correlation is inverted, i.e., the districts that spend next to the most money on education are populated by families having next to the lowest median family incomes while the districts spending the least have the highest median family incomes. It is evident that, even if the conceptual questions were answered favorably to appellees, no factual basis exists upon which to found a claim of comparative wealth discrimination.

⁶¹ Also, it should be recognized that median income statistics may not define with any precision the status of individual families within any given district. A more dependable showing of comparative wealth discrimination would also examine factors such as the average income, the mode, and the concentration of poor families in any district.

*From the Dissenting Opinion of Justice Marshall
at 411 U.S. 1, 94-96*

A number of theories of discrimination have, to be sure, been considered in the course of this litigation. Thus, the District Court found that in Texas the poor and minority group members tend to live in property-poor districts, suggesting discrimination on the basis of both personal wealth and race. See F. Supp., at 282 and n. 8. The Court goes to great lengths to discredit the data upon which the District Court relied, and thereby its conclusion that poor people live in property-poor districts.⁵⁵ Although I have serious doubts as to the correctness of

⁵⁵ I assume the Court would lodge the same criticism against the validity of the finding of a correlation between poor districts and racial minorities.

the Court's analysis in rejecting the data submitted below,⁵⁶ I have no need to join issue on these factual disputes.

I believe it is sufficient that the overarching form of discrimination in this case is between the school children of Texas on the basis of the taxable property wealth of the districts in which they happen to live.

⁵⁶ The Court rejects the District Court's finding of a correlation between poor people and poor districts with the assertion that "there is reason to believe that the poorest families are not necessarily clustered in the poorest property districts" in Texas (*Ante*, at 23). In support of its conclusion the Court offers absolutely no data—which it cannot on this record—concerning the distribution of poor people in Texas to refute the data introduced below by appellees; it relies instead on a recent law review note concerned solely with the State of Connecticut, Note, A Statistical Analysis of the School Finance Decisions: On Winning Battles and Losing Wars, 81 Yale L. J. 1303 (1972). Common sense suggests that the basis for drawing a demographic conclusion with respect to a geographically large, urban-rural, industrial-agricultural State such as Texas from a geographically small, densely populated, highly industrialized State such as Connecticut is doubtful at best.

Furthermore, the article upon which the Court relies to discredit the statistical procedures employed by Professor Berke to establish the correlation between poor people and poor districts, see n. 11, *supra*, based its criticism primarily on the fact that only four of the 110 districts studied were in the lowest categories, which were determined by property per pupil, and most districts clustered in the middle three groups. See Goldstein, Interdistrict Inequalities in School Financing: A Critical Analysis of *Serrano v. Priest* and Its Progeny, 120 U. Pa. L. Rev. 504, 524 n. 67 (1972). See also *ante*, at 26-27. But the Court fails to note that the four poorest districts in the sample had over 50,000 students which constituted 10% of the students in the entire sample. It appears, moreover, that even when the richest and the poorest categories are enlarged to include in each category 20% of the students in the sample, the correlation between district and individual wealth holds true. See Brief for the Governors of Minnesota, Maine, South Dakota, Wisconsin, and Michigan as *amici curiae* 17 n. 21.

Finally, it cannot be ignored that the data introduced by appellees went unchallenged in the District Court. The majority's willingness to permit appellants to litigate the correctness of those data for the first time before this tribunal—where effective response by appellees is impossible—is both unfair and judicially unsound.

REFERENCES

- Aaron, H. J. *Who pays the property tax? A new view*. Washington, D.C.: Brookings Institution, 1975.
- Bahl, R. Studies on determinants of public expenditures: A review. In S. J. Mushkin & J. F. Cotton (Eds.), *Functional federalism: Grants-in-aid and PPB systems*. Washington, D.C.: George Washington University, State-Local Finances Project, 1968.
- Benson, C. S. *The economics of public education* (2nd ed.). Boston: Houghton-Mifflin, 1961.
- Benson, C. S. *Perspectives on the economics of education*. Boston: Houghton-Mifflin, 1963.
- Benson, C. S. *Final report to the Senate Select Committee on School District Finance*. Sacramento, Calif.: California Office of State Printing, 1972.
- Berke, J. S. *Answers to inequity: An analysis of the new school finance*. Berkeley, Calif.: McCutchan, 1974.
- Berke, J. S., Carnevale, D., Morgan, R., & White, R. The Texas school finance case: A wrong in search of a remedy. *Journal of Law and Education*, 1972, 1(4), 659-686.
- Berke, J. S., Kirst, M. W., & Usdan, M. D. *The new era of state education politics: The case of school finance*. Manuscript submitted for publication, 1977.
- Berke, J. S., & Moskowitz, J. H. Research in school finance: A concern with equity and equality. In L. S. Shulman (Ed.), *1976 Review of Research in Education 4*. Itasca, Ill.: Peacock, 1977.
- Burkhead, J. *Public school finance: Economics and politics*. Syracuse, N.Y.: Syracuse University Press, 1964.
- Burkhead, J., Fox, T. G., & Holland, J. W. *Input and output in large city high schools*. Syracuse, N.Y.: Syracuse University Press, 1967.
- Burkhead, J., & Miner, J. *Public expenditures*. Chicago: Aldine-Atherton, 1971.
- Campbell, A. K., & Sacks, S. *Metropolitan America: Fiscal patterns and governmental systems*. New York: Free Press, 1967.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. *Equality of educational opportunity*. Washington, D.C.: U.S. Department of Health, Education, and Welfare, Government Printing Office, 1966.
- Commission to Study School Finance and Equal Educational Opportunity. *Financing Connecticut's schools* (Final Report). Hartford, Conn.: Author, 1975.
- Coons, J. C., Clune, W. H., III, & Sugarman, S. D. Educational opportunity: A workable constitutional test for state finance structures. *California Law Review*, 1969, 57, 305.
- Coons, J. C., Clune, W. H., III, & Sugarman, S. D. *Private wealth and public education*. Cambridge, Mass.: Harvard University, Belknap Press, 1970.
- Cubberley, E. P. *School funds and their apportionment*. New York: Columbia University, Teachers College, 1906.
- Goertz, M. E., & Moskowitz, J. H. *Plain talk about school finance*. Princeton, N.J.: Education Policy Research Institute of the Educational Testing Service, 1976. (Monograph)
- Goldstein, S. Interdistrict inequalities in school financing: A critical analysis of *Serrano v. Priest* and its progeny. *University of Pennsylvania Law Review*, 1972, 120, 504, 524.

- Governor's Citizens' Committee on Education. *Improving education in Florida* (Report). Tallahassee: Author, 1973.
- Guthrie, J. W., Kleindorfer, G. B., Levin, H. M., & Stout, R. T. *Schools and inequality*. Cambridge, Mass.: MIT Press, 1971.
- Hickrod, G. A., Hubbard, J. B., & Yang, W. C. *1973 reform of the Illinois general purpose grant-in-aid: A description and an evaluation*. Unpublished manuscript, University of Illinois, Department of Administration, no date.
- Hirsch, W. *The economics of state and local government*. New York: McGraw-Hill, 1970.
- James, H. T., Kelly, J. A., & Garms, W. I. *Determinants of educational expenditures in large cities of the United States*. Stanford, Calif.: Stanford University, School of Education, 1966.
- Ladd, H. F. Local education expenditures, fiscal capacity, and the composition of the property tax base. *National Tax Journal*, 1975, **28**, 145-158.
- Lawyers' Committee for Civil Rights Under Law. School finance litigation: A strategy session. *Yale Review of Law and Social Action*, 1971, **2**, 2.
- Lennon, R. T. Testimony as expert witness on psychological testing. Unpublished manuscript, Harcourt, Brace, & World, New York, no date.
- Levin, B., Muller, T., Scanlon, W., & Cohen, M. *Public school finance: Present disparities and fiscal alternatives* (Report prepared for the President's Commission on School Finance). Washington, D.C.: Urban Institute, 1972.
- Levin, H. M. *Education, life chances, and the courts: The role of social science evidence*. Unpublished manuscript, Stanford University, 1976.
- Lindquist, R. E., & Wise, A. E. Developments in education litigation: Equal protection. *Journal of Law and Education*, 1976, **5**, 1.
- Morrison, H. *School revenue*. Chicago: University of Chicago Press, 1930.
- National Education Finance Project. *Future directions for school financing: A response to demands for fiscal equity in American education*. Gainesville, Fla.: Author, 1971.
- Netzer, R. *The economics of the property tax*. Washington, D.C.: Brookings Institute, 1966.
- New York State Commission on the Quality, Cost, and Financing of Elementary and Secondary Education. *The Fleischmann Report on the quality, cost, and financing of elementary and secondary education in New York State*. New York: Viking Press, 1973.
- Odden, A., & Vincent, P. *Analysis of the school finance and tax structure of Missouri: Background research of the educational finance cases of the Governor's Conference on Education*. Denver: ECS Education Finance Study, 1976.
- Rivlin, A. Forensic social science. *Harvard Educational Review*, 1973, **43**, 61-75.
- Rosen, P. L. *The Supreme Court and social science*. Urbana: University of Illinois Press, 1972.
- Sacks, S. *City schools/suburban schools: A history of fiscal conflict*. Syracuse, N.Y.: Syracuse University Press, 1972.
- Shalala, D. E., & Williams, M. F. Political perspectives on recent efforts to reform school finance. In S. K. Gove & F. M. Wirt (Eds.), *Political science and school finance*. Lexington, Mass.: Heath, 1976. (Reprinted from *Policy Studies Journal*, 1976, **5**, 367-375.)
- Sommers, A. A., & Wolf, B. L. *Disaggregation and analyzing educational equality issues: Methods and results*. Philadelphia: Federal Reserve Bank of Philadelphia, 1975.
- A statistical analysis of the school finance decisions: On winning battles and losing wars. *Yale Law Journal*, 1972, **81**, 1303.

- Strayer, G. D., & Haig, R. M. *Financing of education in the state of New York*. New York: Macmillan, 1923.
- Thomas, J. A. *The productive school: A systems analysis approach to educational administration*. New York: Wiley, 1971.
- Updegraff, H., & King, L. A. *Survey of the fiscal politics of the state of Pennsylvania in the field of education*. Philadelphia: University of Pennsylvania, 1922.
- U.S. Advisory Commission on Intergovernmental Relations. *Fiscal balance in the American federal system* (Vol. 1). Washington, D.C.: Government Printing Office, 1967.
- Wiley, D. E., & Harnischfeger, A. Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher*, 1974, 3(4), 7-12.
- Wise, A. E. Is denial of equal opportunity constitutional? *Administrators Notebook*, 1965, 10(6), 1-4.
- Wise, A. E. *Rich schools, poor schools: The promise of equal educational opportunity*. Chicago: University of Chicago Press, 1968.
- Wise, A. E. The courts and social science. Unpublished manuscript, Education Policy Research Institute, Washington, D.C., 1977.
- Wolf, E. P. Social science and the courts: The Detroit case. *Public Interest*, 1976, 24.

NAME INDEX

Page numbers in italics refer to the reference lists at the ends of the chapters.

- Aaron, H. J., 587, 643
 Acland, H., 100
 Adcock, C., 302, 317
 Adkins, D. C., 32, 36, 104
 Adler, A., 319, 322, 354, 369, 377
 Aebli, H., 292, 312
 Agard, F., 449, 465
 Agnes, Sister, 173, 183
 Aid, F. M., 428, 465
 Alexander, 618
 Alexander, W. P., 74, 94
 Allen, D. W., 247, 263
 Allen, J. P. B., 409, 421, 423-424, 428, 455, 465
 Allport, G. W., 513, 517
 Almy, M., 319, 346, 349, 388
 Alteneder, L. E., 360-361, 389
 Anastasi, A., 27, 94
 Anderson, G. L., 260
 Anderson, J. D., 343, 388
 Anderson, J. E., 89, 94, 346, 348-349, 391
 Anderson, M. A., 160, 177, 181, 183
 Anderson, R. C., 300, 312
 Anderson, S., 376
 Andersson, T., 440, 465
 Anderzhon, M. L., 175, 183
 Andrews, D. L., 173, 183
 Angoff, W. H., 44, 47, 68-70, 77-78, 94
 Anisfeld, M., 436, 442, 465
 Apostel, L., 307, 312
 Arbib, A., 285, 313
 Argyris, C., 502, 317
 Aristotle, 334
 Asbell, A. E., 132, 193
 Ash, P., 80, 95
 Asher, J., 441, 465
 Athey, I. J., 299, 313
 Auden, W. H., 378, 388
 Austin, G., 274, 313
 Austin, J. L., 428, 465
 Averill, L. A., 349, 365-366, 388
 Ayer, A. M., 131, 183
 Ayer, F. C., 135, 161, 183
 Ayllon, T., 233-235, 238, 260
 Ayres, L. P., 114, 117-118, 125, 140, 146, 183
 Azrin, N. H., 204, 208, 233-235, 238, 260
- Babbitt, E. H., 110, 123, 183
 Bach, W. G., 380, 388
 Bachmann, H. M., 152, 183
 Bagster-Collins, E., 123-124
 Bahl, R., 587, 643
 Bailey, N., 122, 445, 465
 Baker, F. M., 149, 183
 Baker, K. H., 48, 95
 Bakke, E. W., 502, 517
 Banathv, B., 419, 465
 Bandura, A., 242, 244, 246-249, 252, 260
- Bane, M. J., 100
 Barbanel, L., 438, 468
 Barik, H. C., 436, 441, 449, 470, 474
 Barnard, C., 502, 504, 506, 517
 Barnhart, C. L., 123, 184
 Barrie, W. B., 426, 465
 Barth, R. S., 495-496, 517
 Barton, A. H., 147, 184
 Bateman, 607
 Baxter, B., 80, 95
 Bechtoldt, 66
 Beck, J. M., 153, 187
 Becker, W. C., 233, 238, 264
 Bedillon, J. C., 131
 Beers, C. W., 358, 389
 Beilin, H., 297, 313
 Beitell, K. R., 170, 184
 Bellugi, U., 113, 184
 Bennett, G. K., 41, 65, 95
 Bennett, H. E., 117, 184
 Benson, C. E., 360-361, 389
 Benson, C. S., 586, 618, 643
 Berberich, J. P., 246, 263
 Berdie, 66
 Bereiter, C., 229, 260
 Berke, J. S., 581, 583, 594-595, 618, 621, 640, 642, 643
 Berlyne, D. E., 272-273, 313, 494, 517
 Bernard, H. W., 359, 362, 365-366, 389
 Bernfeld, S. B., 336, 373, 389
 Bernstein, I. D., 330, 389
 Berry, A., 168, 184
 Bessel, F. W., 5
 Beth, E. W., 272, 278, 313
 Bettelheim, B., 319-320, 331, 389
 Betts, E. A., 135, 163, 181, 184
 Beust, E., 160, 184
 Bever, T. G., 85, 101
 Bexton, W. H., 494, 517
 Bigelow, M. A., 373, 389
 Bijou, S. W., 204, 233, 238, 260
 Binet, A., 7-12, 14-16, 27, 30, 48, 95
 Bingham, 8
 Birkmaier, E. M., 457, 465
 Birnbaum, A., 48, 95
 Birnbrauer, J. S., 233, 260
 Bjorklund, E., 530, 579
 Blackmun, H. A., 583
 Blair, G. M., 350, 355-356, 389
 Blanchard, P., 332, 366, 389, 391
 Bloch, B., 404, 465
 Block, J. H., 229, 261
 Bloom, B. S., 89-90, 95, 229, 261
 Bloomfield, L., 108, 146, 184, 404-405, 410, 412, 458, 465-466
 Boalt, G., 527, 551-552, 577
 Boas, G., 111, 184

- Böcher, M., 23, 95
 Bock, R. D., 50, 95
 Bolinger, D. L., 413, 416-417, 420-421, 426, 466
 Bolton, F. B., 155, 184
 Bolvin, J. O., 228, 263
 Bond, G. L., 129, 151, 153, 181, 188
 Bonham, D. N., 147, 193
 Bono, S. F., 225, 264
 Boring, E. G., 12, 95
 Bormuth, J. R., 160, 179, 184
 Boud, D. J., 227, 261
 Bovet, M., 278, 297, 314
 Bowen, J. D., 418, 474
 Bower, G., 247, 250, 261-262
 Boynton, P. L., 346, 349, 389
 Brandeis, L. D., 608
 Bravais, A., 20
 Brazer, H., 587
 Breckenridge, M. E., 346, 349, 389
 Breed, F., 181
 Bregman, E. O., 10, 103
 Breitwieser, J. V., 351, 389
 Brennan, W. J., Jr., 583, 596, 598
 Bressler, M., 581
 Brest, 607
 Bridge, W. A., 227, 261
 Bridges, E. M., 498, 517
 Bridgman, P. W., 59, 95
 Brigham, C. C., 9, 42-43, 67-68, 95
 Brill, A. A., 321
 Brim, O. G., 80-81, 95
 Brinkman, A. P., 174, 184
 Broadbent, D. E., 253-254, 261
 Brody, E. B., 85, 95
 Brody, N., 85, 95
 Bromsjö, B., 556, 577
 Brooks, F. D., 346, 349, 389
 Brooks, N., 414, 457, 466
 Brown, 607
 Brown, B. S., 230, 265
 Brown, H. D., 427, 466
 Brown, R., 131, 445, 466
 Brown, R. M., 143, 184
 Brown, T. G., 428, 466
 Brown, W., 10, 21, 46, 95
 Bruce, W. F., 351, 353, 356, 389
 Brudner, H. J., 228, 261
 Bruner, J. S., 274, 277, 280, 313, 491, 520
 Bryan, M. M., 71, 98
 Bryson, L., 181, 184
 Buchanan, M. A., 148, 184, 457, 466
 Buckingham, B. R., 11, 114, 126, 135, 163-164, 181, 184
 Burger, W., 583
 Burgess, C., 111, 196
 Burk, A., 586
 Burkhead, J., 586-588, 643
 Burks, B. S., 481, 517
 Burnham, J. C., 319, 321, 389
 Burnham, W. H., 359-361, 366, 389
 Buros, O. K., 2, 27, 30-32, 38-40, 44, 64-69, 95, 153-155, 184
 Burstall, C., 432, 436-437, 441, 449, 466
 Burt, C., 51, 73, 75, 82, 96
 Burt, M. K., 443, 445, 468
 Bush, R., 274, 314
 Buswell, G. T., 131, 143, 158, 165, 173, 181, 184
 Cabell, J. B., 376
 Caffrey, J., 55, 100
 Cameron, R., 250, 263
 Campbell, A. K., 587, 643
 Campbell, D. T., 62, 97
 Campbell, E. Q., 588, 643
 Campbell, N. R., 59, 96
 Campbell, R. F., 498-499, 503, 505-506, 517-518
 Cancro, R., 85, 96
 Carmichael, K., 225, 262
 Carnevale, D., 595, 640, 643
 Carpenter, H. M., 123, 174, 176, 185
 Carroll, H. A., 366, 389
 Carroll, J. B., 1, 49, 55-59, 66-67, 86-87, 90, 96-97, 123, 126-127, 141, 185, 255-256, 261, 400, 405-406, 421, 435-436, 440, 442-444, 449, 458, 462, 466-467, 482, 517
 Carton, A. S., 450, 467
 Cattell, J. M., 6-7, 9-10, 12, 16-17, 24, 30, 42, 57-58, 85, 90, 96, 99
 Cazden, C. B., 246, 261
 Cecile, Sister, 168, 173, 187
 Cellerier, G., 307, 313
 Chafe, W., 427, 467
 Chall, J. B., 110-111, 113, 120, 122, 132, 147, 153, 159-161, 163, 167-168, 175, 179, 181, 185-186, 192
 Challman, R. C., 351, 391
 Chambers, W. G., 114, 185
 Chandrasekhar, S., 508, 518
 Charters, W. W., 164, 185
 Chase, H. W., 342, 389
 Chase, S. E., 167, 185
 Chastain, K. D., 420, 422, 449, 457, 467
 Chauncey, H., 80, 96, 114, 185
 Checkovitz, A., 171, 196
 Chester, R., 135, 185
 Childs, H. G., 110, 196
 Chomsky, N., 274, 278, 313, 400, 406, 408, 414-419, 421, 430, 443, 458, 467
 Christie, T., 75, 96
 Ciruti, J. E., 415, 466
 Clark, B. R., 116, 185
 Clark, D. L., 478, 487, 502, 518-519
 Clark, H. H., 319
 Clark, W. W., 28, 30, 103
 Clement, J. A., 162, 185
 Clifford, G. J., 107, 109, 116, 148, 173, 182, 185; *see also* Joncich, G.
 Clifford, J. C., 399-400, 467
 Clowes, R. M., 603
 Clune, W. H., III, 589-590, 597, 643
 Cobb, M. V., 10, 103
 Coffman, W. E., 69, 102
 Cohen, D., 100
 Cohen, M., 618, 644
 Cohen, M. R., 511, 518
 Cohen, S., 432, 466
 Coladarsi, A. P., 502, 509, 512-514, 518
 Cole, L., 108, 131-132, 150, 157, 165, 167-168, 170, 181, 185; *see also* Pressey, L. C.
 Cole, L. E., 351, 353, 389
 Coleman, A., 116, 128, 135, 138-139, 143, 149, 168-169, 185, 196
 Coleman, J. S., 588, 600-601, 610, 643
 Coleman, W. H., 181
 Columba, Sister, 173, 185

- Colvin, C. M., 114, 186
 Colvin, S. S., 484, 518
 Comenius, J. A., 224
 Commins, W. D., 351, 353, 355, 389
 Comrey, A. L., 436, 472
 Conant, J. B., 43, 72, 96, 511, 518, 543, 551, 577
 Conger, J. J., 346, 349, 393
 Connelly, F. M., 397-398, 473
 Connolly, P. C., 426, 467
 Cook, W. W., 114, 186
 Cooke, D., 451, 467
 Cooley, W. W., 41, 62, 97-98
 Coombs, C. H., 56, 97
 Coons, J. C., 588-590, 597, 605, 612-613, 643
 Cooper, K. S., 175, 186
 Cooper, R., 432, 468
 Copeland, R. W., 301, 313
 Corbally, J. E., 498, 505, 517
 Corder, S. P., 408, 422-424, 430, 445-446, 455, 465, 467
 Cortat, I. H., 374, 389
 Cornelius, V., 139, 197
 Cornell, E., 138, 186
 Cottle, W. C., 153, 188
 Cowan, S., 436, 467
 Coy, G. L., 181
 Crano, W. D., 62, 97
 Crape, L., 175, 190
 Crawford, A. B., 36, 97
 Cremin, L. A., 5, 97, 113, 116, 186, 319, 371, 389, 490, 518
 Criper, C., 429, 467
 Crocker, L., 139, 186
 Cronbach, L. J., 47-48, 79, 84, 97, 121, 134, 157, 173, 186, 319, 351, 353, 355, 389, 480, 518
 Crothers, E., 450, 467
 Cruze, 356
 Csikszentmihalyi, M., 480, 508, 518
 Cubberley, E. P., 498, 500, 518, 583-585, 619, 643
 Cuff, N. B., 351, 353, 355, 357, 389-390
 Cureton, E. E., 153-155, 186, 191
 Curran, C. A., 438, 468
 Curtis, D., 175, 186
 Curtis, F. D., 125, 130-131, 142, 174, 181, 186
 Cutts, N. E., 365-366, 390

 Dahlöf, U., 556, 558, 571, 573, 577-578
 Dahrendorf, R., 551, 577
 Failey, J. T., 41, 98
 Dale, E., 108, 110, 113-115, 135-136, 153, 159, 179, 181, 186
 Dalton, J., 378
 Danaher, B. G., 250, 263
 Danziger, M., 126, 186
 Darwin, C. R., 5, 369, 375, 378
 Davidson, W. M., 86, 97
 Davies, G. F., 116, 189
 Davies, P., 123, 185
 Davis, F. B., 70-71, 154-155, 181, 187
 DeAngelo, R. W., 171, 187
 Dearborn, F. R., 150, 187
 DeLima, A., 372, 390
 Dell'orto, K. M., 424, 473
 Dell'orto, V. J., 424, 473
 Derman, L., 297, 305, 314, 316
 deVilliers, J., 445, 468
 deVilliers, P., 445, 468

 DeVoto, B., 378, 390
 DeVries, R., 302-303, 305-306, 313-314
 Dewey, J., 108, 118, 187, 199, 201-202, 261, 320, 371, 374, 381-385, 388, 390, 458, 490, 514-515, 518
 Dewey, J. C., 127, 187
 Dickens, C., 172
 Dickerson, D. F., 135, 187
 Dickson, W. J., 499, 520
 Diller, K. C., 420, 443, 468
 Dillon, J. T., 480, 518
 DiPietro, R., 418, 468
 Disick, R. S., 438, 468
 Dixon, N., 139, 197
 Doar, J., 613
 Dobbin, J. E., 80, 96, 114, 185
 Docter, R., 81, 99
 Dodge, R. E., 175, 187
 Dolch, E. W., 113-114, 126, 135-136, 145, 147, 149-150, 162, 164, 167, 181, 184-185, 187
 Dollard, J., 244, 263, 343, 390
 Donnelly, H. E., 174, 187
 Doolittle, M. H., 61, 97
 Doran, E. W., 108, 110, 123, 187
 Dorothy, Sister, 168, 173, 187
 Douglas, O. B., 355, 390
 Douglas, W. O., 583, 596, 598
 Downey, M. T., 42-43, 97
 Drabman, R., 239, 264
 DuBois, P. H., 5, 37, 49, 68-69, 97, 101
 Dulay, H. C., 443, 445, 468
 Dunkel, H. B., 449, 463
 Dunlap, J. W., 37, 97
 Dunlap, K., 323
 Durning, K., 247, 264
 Durrell, D., 122, 134, 166, 181, 187
 Dutton, W. H., 342, 390
 Dvorak, B. J., 41, 97

 Easley, H., 27
 Ebbinghaus, H., 6-7, 97
 Eckerson, L. D., 154, 193
 Edenman, R., 532, 535
 Edgerton, H. A., 57, 97
 Edgerton, R. B., 177, 187
 Edmonson, J. B., 160, 176, 195
 Edwards, A. L., 491, 518
 Einstein, A., 376-377, 507, 518
 Ekman, G., 537-540, 577
 Ekstein, R., 320, 390
 Ekstrom, R. B., 56, 59, 97-99
 Elam, S. K., 132, 194
 Eldridge, R. C., 124
 Elkind, D., 288, 299, 313
 Elkins, A. D., 175, 187
 Elliott, C., 296, 313
 Ellis, D. B., 362, 390
 Ellis, R. S., 351, 353, 390
 Ellison, M. L., 57, 97
 Elmgren, J., 526, 545, 577
 Emme, E. E., 355, 390
 Engelmann, S., 229, 260, 297, 313
 Englehart, M. D., 153, 187
 Erikson, E., 325, 337-339, 373, 390; *see also* Homburger, E.
 Ervin-Tripp, S. M., 436, 445, 468
 Estes, W. K., 213, 255, 261

- Evans, B., 162, 171-172, 187, 192
 Evans, G. K., 170, 191
 Evenden, E. S., 148, 158, 187
 Eysenck, H. J., 82, 86-87, 97
- Faber, C. F., 504, 518
 Fagin, B., 351, 353, 389
 Falick, M. L., 365, 390
 Faraday, M., 508
 Farrand, L., 6, 96
 Farrar, C. R., 116, 189
 Fasnacht, C., 240, 263
 Faucett, L., 138-139, 146, 187
 Faust, G. W., 300, 312
 Fea, H. R., 110, 194
 Featherstone, W. B., 35, 101
 Fechner, G., 5, 15, 97
 Feenstra, H. J., 436-437, 468-469
 Feigenbaum, E. A., 274, 313
 Feldman, J. G., 274, 313
 Fenichel, O., 328, 390
 Fennell, E. L., 121, 150, 164, 187
 Fenton, N., 366, 390
 Ferguson, G., 48, 58, 97
 Ferster, C. B., 204, 207, 215-216, 261
 Feuer, L. S., 383, 390
 Feurzeig, W., 307, 313
 Feynman, R. P., 507, 518
 Fieldstra, C., 142
 Fife, R. H., 139, 187
 Figurel, J. A., 171, 187
 Filley, M. E., 154, 193
 Fillmore, C., 427, 468
 Fincher, C., 77, 97
 Findley, W. G., 71, 98
 Finney, D. J., 48, 98
 Fisher, R. A., 19-20, 48, 98
 Fishman, J. A., 432-433, 468
 Fitzgerald, J. A., 181
 Flamer, G. B., 85, 99, 298, 314
 Flanagan, J. C., 22, 40-41, 98, 228, 261
 Flavell, J. H., 268-270, 273, 277-279, 293, 295, 313
 Fleckeisen, 124
 Flesch, R., 126, 181, 187
 Fodor, J., 426, 470
 Fogelman, K. R., 296, 313
 Foran, T. G., 145, 164, 173, 187
 Forbes, F. W., 153, 188
 Foró, M., 298, 314
 Ford, M. P., 85, 99
 Fox, M. J., 401, 453-454, 468
 Fox, T. G., 588, 643
 Francis, W. N., 127, 138, 141, 146-147, 188, 191
 Franklin, B., 130
 Franzen, R. H., 116, 159, 188
 Frederiksen, 66
 Freeman, 356
 French, J. L., 25, 63, 98
 French, J. W., 56, 59, 98-99
 Freud, A., 319-320, 325-327, 339, 353, 390
 Freud, S., 319-324, 331, 334-336, 342-343, 348-350, 354-355, 368-385, 387, 390, 458, 501
 Friedman, K. C., 174, 188
 Friedman, R. J., 449, 474
 Fries, C. C., 109, 124, 137-138, 144, 154, 160, 188, 406, 458, 460, 468-469
 Frohlich, M., 447, 471
- Frost, N., 59, 62, 100
 Fulker, D. W., 88, 98
 Funchess, L. W., 170, 188
 Furby, L., 47, 97
 Furneaux, 87
 Furth, H. G., 278, 302-304, 306, 313
- Gabler, E. R., 175, 188
 Gage, N. L., 319
 Galanter, E., 273-274, 314
 Galileo, 378
 Galton, F., 4-7, 9-10, 14, 16-21, 24, 90, 98
 Gansl, I., 110, 154, 188
 Garcia, R., 441, 465
 Gardner, R. C., 435-438, 468-470
 Garms, W. L., 588, 618, 644
 Garrison, K. C., 114, 132, 173, 188, 351, 353, 390
 Garrison, S. C., 114, 132, 173, 188
 Gates, A. I., 124, 129, 136, 151, 160, 161, 188-332, 351, 353-357, 390-391
 Gauss, K. F., 20
 Gaylord, R. H., 49, 105
 Gentner, D. R., 256, 263
 Georges, J. S., 173, 188
 Gerberich, J. R., 189
 Gesell, A. L., 348-349, 391, 439, 469
 Gesell, B. C., 349, 391
 Getzels, J. W., 480, 492-493, 495, 499-500, 502-503, 506, 508-509, 512-514, 518
 Gilbert, C., 171, 196
 Gillette, G. M., 181
 Gillette, J. M., 188
 Gimson, A. C., 402, 469
 Ginsburg, H., 268, 270, 273, 295, 313
 Gintis, H., 100
 Gjerde, C. M., 363, 394
 Gladstone, W. E., 508
 Glaser, R., 199, 213, 223, 228, 251, 257, 261, 263
 Glass, D. C., 81, 95
 Gleason, H. A., Jr., 427, 429, 469
 Gleser, G. C., 48-49, 97, 101
 Goddard, H. H., 7-8, 98
 Goertz, M. E., 590, 643
 Goettel, 618
 Goldberg, I., 41, 98
 Goldiamond, I., 210, 263
 Goldstein, S., 595, 607, 640, 642, 643
 Goodenough, F. L., 346, 348-349, 391
 Goodman, J., 251, 263
 Goodnow, J., 274, 313
 Gorbet, F., 443, 469
 Gordon, B., 398, 438, 470
 Gordon, E. W., 78, 98
 Gorham, W. A., 41, 98
 Gorman, F. H., 173, 188
 Goslin, D. A., 80-81, 98
 Gougenheim, G., 411, 469
 Gouin, F., 402, 459, 469
 Grahm, A., 573, 577
 Grauberg, W., 425, 469
 Gray, S., 351, 353, 390
 Gray, W. S., 110-111, 119-120, 142, 160, 181, 188
 Gréco, P., 307, 314
 Green, B. F., 62, 98
 Green, D. R., 85, 98, 298, 314
 Green, H. H., 170, 188
 Greenbaum, S., 423, 473
 Greene, H. A., 135, 142, 147, 189

- Greenfield, P. M., 280, 313
 Greeno, J. G., 256, 261
 Gregg, R. T., 503, 517
 Gregory, C. A., 143, 189
 Griffin, A., 76, 96
 Griffiths, D. E., 498, 503, 518
 Grinstead, W. J., 123, 189
 Grittner, F. M., 457, 469
 Grize, J. B., 307, 312, 314
 Groen, G., 267
 Gross, R. E., 177, 189
 Groves, E. R., 366, 391
 Gruenberg, S. M., 380
 Guba, E. G., 478, 503, 518
 Guilford, J. P., 27, 40-41, 56-57, 83, 85, 99, 154, 189, 486
 Gulliksen, H., 44, 46, 64, 99
 Gustafson, L. M., 166, 191
 Guthrie, E. R., 351, 353, 391
 Guthrie, J. W., 618, 644
 Guttman, L., 18, 49, 55, 64, 99
 Gwynn, J. M., 171, 173, 189

 Hagen, E., 41, 103, 155, 197
 Haggerty, M. E., 152, 189
 Haig, R. M., 585, 645
 Hakstian, A. R., 42, 99
 Hakuta, K., 445, 469
 Hale, N. G., Jr., 319, 374, 391
 Hall, G. S., 111, 113, 189, 199, 321-323, 340, 342, 349, 391
 Hall-Quest, A. L., 157-158, 189, 351, 355, 391
 Halliday, M. A. K., 407, 411, 427, 429, 457, 469
 Halpern, H., 126, 186
 Halpin, A. W., 503, 519
 Hamilton, M. H., 149, 189
 Hancock, C. R., 435, 469
 Hanley, E. M., 235, 261
 Hanna, P. R., 116, 189
 Hanratty, M., 247, 263
 Hansen-Bede, L., 445, 469
 Hapiewicz, W. G., 236, 262
 Harap, H., 159, 189
 Harbeson, C., 175, 189
 Hargreaves, M., 432, 466
 Harley, B., 397, 449, 474
 Harlow, H. F., 519
 Harman, H. H., 56, 99
 Harnischfeger, A., 82, 99, 588, 645
 Härnqvist, K., 527, 538-541, 573, 577
 Harris, H. I., 329, 391
 Harris, M. B., 247, 260
 Hartmann, G. W., 110, 189
 Hartmann, H., 325, 329, 335, 356, 387, 391
 Hatch, E., 445, 469
 Hausman, J., 170, 189
 Hayes, E. J., 78, 102
 Hayes, J. R., 256, 262, 264
 Healy, W., 343, 391
 Hebb, D. O., 487, 494, 519
 Hébrard, P., 445, 471
 Hecht, B., 376
 Heddens, J. W., 178, 195
 Heidbreder, E., 324, 391
 Helmholtz, H., 14, 99
 Hendricks, W. O., 429, 469
 Hendrickson, G., 170, 189
 Henmon, V. A. C., 11, 25-26, 99, 148, 169, 181, 189
 Henry, N. B., 204, 262, 498, 503, 519
 Herbart, J. F., 350
 Heron, W., 494, 517
 Herrnstein, R. I., 81, 99
 Hertzberg, O. E., 357, 391
 Hewett, F., 233, 262
 Hewitt, T. B., 168, 189
 Heyns, B., 100
 Hickrod, G. A., 618, 644
 Hildreth, G., 135-136, 145, 158, 168, 181, 189
 Hilgard, E. R., 204, 247, 262, 319, 324, 329, 354, 391, 493, 519
 Hill, J. C., 531, 391
 Hill, J. H., 247, 263
 Hilliard, G. H., 119, 190
 Hillis, C. B., 343
 Hills, J., 504, 519
 Hills, J. R., 80, 99
 Hirsch, W., 587, 644
 Hobson, C. J., 588, 643
 Hobson, C. V., 363-364, 391
 Hoch, A., 321
 Hockett, J. A., 164, 168, 181-182, 190
 Hoepfner, R., 41, 56, 99
 Hoffer, W., 320, 391
 Hoffman, F. J., 374, 391
 Hofstadter, R., 5, 99, 374, 391
 Hölder, L. O., 14, 59, 99
 Holland, J. W., 588, 643
 Hollingworth, H. L., 351, 391
 Holmen, M. G., 81, 99
 Holt, E. B., 323, 392
 Holz, W. C., 208, 260
 Holzinger, K., 13, 23
 Homburger, E., 337, 392; *see also* Erikson, E.
 Homme, L. E., 249, 262
 Hooper, F., 277, 316
 Horn, E., 125-129, 134-135, 137, 140, 142, 147-148, 150, 157, 163, 169, 181, 190
 Horn, J. L., 42, 57, 99
 Horn, M. D., 127, 157
 Horn, T. D., 150, 190
 Horn, W. O., 42, 99
 Horne, H. H., 351, 392, 484, 519
 Horner, V., 170, 190
 Horney, K., 322, 343, 374
 Horrocks, J. E., 154, 190
 Horst, P., 37, 61, 99
 Hosford, R. E., 247, 265
 Hotelling, H., 51, 61, 99
 Hoyt, C., 48, 100
 Hubbard, J. B., 618, 644
 Hubbard, L. J., 438, 470
 Huey, E. B., 168, 190
 Hugh, D. D., 483, 519
 Hughes, R., 376
 Hull, C. L., 204, 254
 Hull, G. L., 273, 314
 Humphrey, C. W., 169, 190
 Humphreys, L. G., 66, 88, 100
 Hunnicutt, C. W., 168, 190
 Hunt, E., 59, 62, 100
 Hunt, J. M., 58, 100, 277, 314, 493-494, 519
 Hunter, E. C., 365, 392
 Hurlock, E. B., 346, 349-350, 392
 Husén, T., 85, 100, 523, 525, 527, 529, 532-533, 535-539, 542, 547, 550-552, 554, 556, 577-578
 Hymes, D., 430, 470

- Ilig, F. L., 349, 391, 439, 469
 Infeld, L., 507, 518
 Inhelder, B., 269, 276, 278, 296-297, 314, 316
 Irmina, Sister, 143, 190
 Isaacs, N., 288, 314
 Iverson, W. J., 168, 190

 Jackson, B., 543, 578
 Jäger, A. O., 56, 100
 Jakobovits, L. A., 398, 435, 438, 470
 James, H. T., 588, 644
 James, L. B., 175, 190
 James, W., 199, 321-322, 375-376, 392, 483-484
 Jamieson, M., 432, 466
 Jarvia, R. A., 417, 422, 470
 Jastrow, J., 322
 Jaynes, J., 227, 262
 Jelleffe, S. E., 321
 Jencks, C., 41, 100
 Jenkins, F., 346, 349-350, 393
 Jenkins, H. B., 132, 193
 Jensen, A. R., 81-82, 87, 89, 100
 Jensen, A. S., 346, 349, 392
 Jensen, D. W., 481, 517
 Jensen, T. J., 502, 519
 Jensen, L., 177, 190
 Jersild, A. T., 343, 346, 348-349, 351, 355, 391-392
 Jespersen, O., 458
 Johanson, E., 556, 578
 John, L., 131, 158, 184
 Johns, J. L., 138, 149, 190
 Johns, R. L., 505, 520, 586, 618
 Johnson, D. D., 126, 190
 Johnson, M. E., 165, 190
 Johnson, R. H., 153, 190
 Johnston, J. M., 227, 262
 Johnston, M., 149, 190
 Joncich, G., 114, 123-124, 181, 191; *see also*
 Clifford, G. J.
 Jonas, S. R., 438, 470
 Jones, A. N., 170, 191
 Jones, E., 321
 Jones, F. W., 181
 Jones, G. F., 411, 470
 Jones, L. V., 60, 100
 Jones, L. W., 12, 102
 Jones, W. F., 191
 Joos, M., 110, 191
 Jordan, A. M., 351, 353, 356, 392
 Joreskog, K. G., 55, 100
 Jorgensen, A. N., 189
 Judd, C. H., 107, 143, 161, 184, 191, 351, 392
 Jung, C. G., 321, 342, 350, 354, 369, 481, 519
 Jung, S. M., 41, 98
 Jurgen, 376
 Justman, J., 114, 198

 Kaeding, F. W., 124
 Kaestle, C. F., 117, 191
 Kaiser, H. F., 53, 55, 100
 Kamü, C., 297, 302-303, 305-306, 308, 313-314, 316
 Kamin, L. J., 82, 87-88, 91, 100
 Kanfer, F. H., 242-243, 250, 262
 Karner, C. J., 383, 392
 Karplus, R., 485-486, 519
 Katz, B., 366, 395
 Katz, J., 414, 426, 470
 Kavetsky, J., 139, 191
 Kazdin, A. E., 230-231, 262
 Kazin, A., 379, 392
 Kearney, N. C., 168, 191
 Keating, R. F., 449, 462, 470
 Keats, 66-67
 Keislar, E. R., 293, 300, 316
 Keliher, A. V., 34, 100
 Keller, F. S., 214, 225-227, 236, 262
 Kellerman, E., 446, 470
 Kelley, A., 149, 191
 Kelley, T. L., 22, 38, 46, 100, 115-116, 152-153, 155, 181, 191
 Kelly, J. A., 588, 644
 Kelly, L. G., 401, 470
 Kennon, L. H. V., 126, 191
 Kenny, D. A., 62, 97
 Kepner, T., 130, 191
 Kerlinger, F. N., 374, 392
 Kessen, W., 276, 314
 Keynes, J. M., 497, 519
 Khampang, P., 445, 470
 Kidder, J. D., 233, 260
 King, C. B., 143, 149, 169, 196
 King, L. A., 585, 643
 Kiparsky, C., 443, 468
 Kirkpatrick, E. A., 108-110, 115, 123, 191, 346, 349, 392
 Kirkpatrick, J. J., 154-155, 191
 Kirp, D., 612, 613
 Kirst, M. W., 618, 643
 Klahr, D., 255, 262
 Klare, G. R., 109, 111, 124-125, 132, 153, 161, 166, 179, 191
 Klausmeier, H. J., 229, 262
 Klein, E., 332, 392
 Klein, M., 333-334, 392
 Klein, R. D., 236, 262
 Kleindorfer, G. B., 618, 644
 Klezevich, S. J., 505, 519
 Knight, F. B., 116, 159, 181, 188
 Knott, T. A., 174, 191
 Koch, S., 273, 314
 Koegel, R. L., 238, 246, 263
 Koerth, W., 351, 393
 Kohlberg, L., 295, 298, 314
 Kolevzon, E. R., 176, 191
 Krantz, D. H., 60, 100
 Krashen, S., 445, 465
 Krasner, L., 204, 229, 238, 262, 265
 Krasner, M., 238, 262
 Krey, A. C., 116, 153, 155, 191
 Kris, E., 329, 335-336, 391-392
 Krug, E. A., 116, 191
 Krumboltz, H. B., 236-237, 262
 Krumboltz, J. D., 236-237, 244, 262
 Kučera, H., 127, 139, 141, 147, 191
 Kuder, G. F., 46, 100
 Kuhlman, C., 31, 276, 314
 Kuhn, T. S., 178, 191, 280-282, 314, 487, 519
 Kulik, C. L., 225, 227, 262
 Kulik, J. A., 225, 227, 262
 Kunstler, 609

 Labov, W., 80, 100
 Ladd, H. F., 587, 644
 Lado, R., 410, 413-414, 457, 460, 470

- LaForge, P. G., 438, 470
 Lakatos, I., 267-268, 281-283, 292, 314
 Lamarck, J., 328
 Lambert, C. M., 100
 Lambert, W. E., 432, 435-437, 449, 453, 465, 469, 470
 Lamendella, J. T., 419, 470
 Lamke, T. A., 25, 64, 101
 Lange-Eichbaum, W., 481, 519
 Lapkin, S., 449, 474
 Laplace, P. S., 20
 Larendau, M., 296, 316
 Larrabee, H. A., 511, 519
 Larrick, N., 181
 LaRue, D. W., 132, 145, 191, 351, 353, 361, 392
 Lashley, K. S., 322-323, 392
 Laudan, L., 283, 314
 Lavatelli, C., 302-303, 306, 314
 Lavoisier, A. L., 378
 Lawley, D. N., 55, 100
 Learned, W. S., 157, 192
 Leary, B. E., 111, 142, 188
 Lee, D. M., 142, 145, 192
 Lee, J. M., 142, 145, 192
 Leech, G., 423, 473
 Lefever, D. W., 64-65
 Leiman, J. M., 55, 102
 Lennon, R. T., 609, 644
 Léon, M., 413, 470
 Léon, P. R., 402, 470
 Leslie, L. A., 169, 181, 192
 Lessenberry, D. D., 170, 192
 Leuba, J. H., 322
 Levenston, E. A., 423, 432, 470
 Levi, E., 606
 Levi, H. S., 175, 192
 Levin, B., 618, 644
 Levin, H. M., 134, 192, 611, 614, 618, 644
 Levin, L., 449, 470
 Levine, F. M., 240, 263
 Levine, J. M., 491, 519
 Levi-Strauss, 278
 Levitt, M., 365, 383, 390, 392
 Levy, D. M., 332, 392
 Levy, P., 50, 101
 Lewin, K., 482, 492, 500-501, 508-509, 519
 Lewis, E. G., 401, 449, 462, 470-471
 Lewontin, R., 82
 Liebert, R. M., 247, 263
 Linden, J. D., 5, 101
 Linden, K. W., 5, 101
 Lindgren, H. C., 366, 393
 Lindman, E., 586
 Lindquist, E. F., 44
 Lindquist, R. E., 614, 644
 Lindsley, O. R., 204, 260
 Lindvall, C. M., 228, 263
 Linn, R. L., 57, 62, 101, 103
 Lipham, J. M., 499, 518
 Lippitt, R., 482, 492, 500, 519
 Lippmann, W., 374, 393
 Liss, E., 329, 393
 Lively, B. A., 131-133, 179, 192
 Livengood, W. W., 164, 192
 Loban, W., 113, 192
 Lockman, R., 491, 520
 Lodge, G., 123-124, 168, 192
 Lochlin, J. C., 88, 101
 Loevinger, J., 18, 47, 49, 101
 Loewenstein, R. M., 329, 335, 391
 Lohnes, P. R., 62, 97
 Lombroso, C., 481, 519
 Lord, F. M., 44, 46-49, 62, 86, 101
 Loretan, J. O., 76, 101
 Lorge, I., 110, 124, 126, 139, 141-142, 151, 179, 181, 192, 196-197
 Lovaas, O. I., 246, 263
 Loveland, G., 164, 192
 Lovell, K., 276, 300, 314
 Low, B., 342, 393
 Luce, D., 274, 314
 Luce, R. D., 60, 100
 Lukas, G., 307, 313
 Lumsdaine, A. A., 222, 263
 Lumsden, J., 50, 101
 Lunneborg, C., 59, 62, 100
 Lunzer, E. A., 276, 314
 Lynch, J. J., 162, 171-172, 187, 192
 Lyons, J., 421, 425, 471
 Ma, R., 432, 468
 Mabry, J. E., 166, 191
 MacGinitie, W. M., 119, 179, 192
 Mackey, W. F., 407, 432, 457, 471
 MacLane, S., 285, 294, 314
 MacLatchy, J. H., 168, 192
 MacLeod, R. B., 510, 519
 Macnamara, J., 427, 463, 471
 MacPhee, E. D., 457, 466
 Madden, C., 445, 465
 Magnusson, P. M., 351, 393
 Mahoney, M. J., 242-243, 249-250, 263, 265
 Maki, I., 139, 146, 187
 Mallinson, G. G., 130, 160, 174, 181, 192
 Malmquist, C. P., 393
 Malmquist, E., 542, 578
 Maloff, R., 176, 191
 Mann, T., 377
 Mannes, E., 285, 313
 Marjoribanks, K., 87, 101
 Marker, R. W., 228, 261
 Marklund, S., 529-530, 575, 578
 Marland, S. P., 228, 265
 Marshall, T., 583, 595-603, 605, 607, 614, 621, 629, 636, 638, 640-641
 Martin, W. E., 346, 349, 393
 Marx, K., 377
 Massad, C. E., 449, 462, 471
 Masters, E. L., 376
 Matherson, G., 376
 Mathews, C. O., 131, 192
 Mayer, G. R., 236, 265
 Mayer, M., 157, 160, 192
 Mayo, E., 499
 McAulay, J. D. E., 163, 176, 192
 McCall, W. A., 114, 181
 McCarthy, D., 111, 113, 192
 McCaul, R. L., 490, 519
 McCawley, J. D., 414, 471
 McClellan, D., 302, 317
 McClure, W. E., 34, 101
 McConnell, T. R., 351, 391
 McCormick, E. J., 254, 263
 McCullough, C. M., 87, 101
 McDermott, J., 612-613
 McDonald, F. J., 247, 263

- McDonald, R. P., 52, 55, 101
 McGinnis, E., 491, 520
 McGuffey, G., 120-121
 McIntosh, A., 407, 457, 469, 475
 McIntyre, R. D., 436, 472
 McKee, G., 128, 151, 192
 McKee, P., 161-162, 181, 193
 McKillop, A. S., 355, 391
 McKinnon, K. R., 450, 471
 McLean, L. D., 449, 474
 McNeill, D., 447, 471
 McNeill, H., 378, 393
 McPartland, J., 588, 643
 Meek, L. H., 373, 393
 Mehler, J., 85, 101
 Meichenbaum, D. H., 250-251, 263
 Meili, R., 56, 101
 Melton, I. W., 213, 263
 Mendenhall, R. M., 24, 103
 Menninger, K., 377, 393
 Merrill, M. A., 7, 103
 Messing, L., 319
 Metcalf, H. C., 499, 519
 Meyer, A., 321-322
 Meyer, H. B., 319
 Michéa, R., 411, 469
 Michelman, F., 613
 Michelson, S., 100
 Miles, J., 154, 193
 Millard, C. V., 346, 349, 393
 Miller, G. A., 273-274, 314
 Miller, H. A., 117, 193
 Miller, L. W., 362, 390
 Miller, N. E., 244, 263, 343, 390, 511, 513, 519-520
 Miller, T. W. G., 549, 578
 Miller, V., 504, 520
 Mills, 507, 509
 Mills, R. E., 161, 193
 Miner, J., 587, 643
 Mischel, W., 244, 263
 Mitchell, D., 117-118, 193
 Mitchell, J. C., 362, 393
 Moberg, S., 532-533, 535, 551, 578
 Modgil, S., 277, 314
 Monmouth, R. A., 132, 193
 Monroe, W. S., 37, 101
 Montero, H. H., 413, 466
 Mood, A. M., 588, 643
 Moore, B., 614
 Moore, H. T., 68
 Moore, J. E., 170, 193
 Moore, R., 210, 263
 Morgan, C. L., 147, 193
 Morgan, J. J. B., 349, 361, 393
 Morgan, R., 595, 640, 643
 Morphet, E. L., 505, 520
 Morphet, M. V., 171, 193, 198
 Morris, I., 127, 193
 Morrison, H., 585-586, 644
 Morrison, V. M., 139, 156, 174, 191, 193
 Morse, W. H., 207, 263
 Mort, P., 586
 Mort, P. R., 35, 101
 Moseley, N., 366, 390
 Moser, W., 131, 193
 Moskowitz, J. H., 583, 590, 643
 Motto, R. L., 320, 390
 Mougeon, R., 445, 471
 Moulton, W. G., 404-405, 418, 571
 Mowrer, O. H., 108, 193
 Muller, T., 618, 644
 Munsterberg, H., 201, 263, 351, 393
 Murphy, G., 491, 519
 Murra, W. F., 162-163, 193
 Murray, H. A., 324
 Murray, H. J., 229, 264
 Mursell, J. L., 351, 353, 393
 Mussen, P. H., 346, 349, 393
 Myers, G. C., 114, 195
 Nagge, J. W., 346, 349, 393
 Naiman, N., 447, 471
 Nance, J. H., 319
 Nanda, H., 48, 97
 Naumberg, M., 372, 393
 Neeley, D. P., 164, 190
 Neill, A. S., 335
 Nelson, A. H., 159, 193
 Nelson, M. J., 25-26, 64, 92, 101
 Nelson, R. J., 438, 471
 Nemser, W., 446, 471
 Netzer, R., 587, 644
 Neufeld, G. G., 435, 471
 Neuhinger, J., 81, 93
 Newell, A., 274, 315
 Newmark, L., 420, 443, 471
 Newport, J. F., 163, 193
 Newton, I., 280, 369
 Neymark, E., 533-534, 539, 545, 578
 Nichols, R. C., 88, 101
 Nicholson, A., 125, 193
 Nickel, G., 418, 471
 Nida, E. A., 436, 471
 Nilsen, D. L. F., 428, 472
 Nisbet, J. F., 481, 520
 Nixon, R. M., 483
 Noblitt, J. S., 423, 472
 Nolte, K. F., 130, 193
 Norman, D. A., 256, 263
 Norris, W. E., 414, 472
 Norworthy, N., 346, 349, 393
 Norwood, C., 74-75, 101
 Novak, J., 229, 264
 Novick, M. R., 46, 48-49, 101
 Nystrand, R. O., 498, 517
 Oberndorf, C. P., 322, 393
 O'Connor, J., 154, 193
 Odden, A., 618, 644
 O'Dell, S., 236, 264
 Oden, M. H., 481, 520
 Ogden, C. K., 159, 193
 O'Higgins, H., 376
 Ojemann, R., 179, 181, 193
 O'Leary, K. D., 233, 236, 238-239, 264
 O'Leary, S. G., 236, 264
 Oller, J., 431, 472
 Olson, D. R., 426, 472
 Olver, R. R., 280, 313
 O'Neill, E., 376
 Oppen, S., 268, 270, 273, 314
 O'Rourke, 154
 Orr, D. B., 41, 98
 Orring, J., 527, 550, 552, 578
 Osgood, C. E., 406, 433, 472

- O'Shea, M. V., 349, 351, 357, 393
 Osherson, D. N., 315
 Otis, A. S., 7-8, 24, 101, 153, 193
 Otto, W., 108, 121, 135, 157, 185, 193
 Overholser, O. V., 177, 193
 Painter, W. I., 129, 131, 194
 Palmer, H. E., 139, 142, 187, 193, 402, 409, 412, 458-459, 472
 Papert, S., 272, 295, 299, 306-310, 312-315
 Paret, I., 319
 Paret, P., 319
 Park, D. G., 323, 393
 Parker, F., 118
 Parker, W. R., 452, 472
 Parsons, 506
 Partridge, G. E., 349, 393
 Patty, W. W., 129, 131, 194
 Paulston, C. B., 431-432, 443, 451, 472
 Paulston, R. G., 544, 578
 Pavlov, I., 204-205, 223-224, 330, 349
 Pearson, G. H. J., 334, 393
 Pearson, K., 14, 19-21, 28, 61, 90, 102
 Pechstein, L. A., 346, 349-350, 393
 Peck, P., 362, 393
 Peller, L. E., 328-329, 394
 Penfield, W. G., 439, 472
 Pennybacker, H. S., 227, 262
 Percival, W. P., 118, 194
 Perloff, B. F., 246, 263
 Pestalozzi, J. H., 350
 Peters, M., 365, 390
 Petersen, T., 169, 190
 Peterson, F., 321
 Peterson, H. A., 351, 394
 Peterson, J., 5, 9, 102
 Phillips, D. C., 319
 Phillips, J. L., 273, 299, 315
 Phillips, J. S., 242-243, 250, 262
 Phillips, M. V., 175, 194
 Piaget, J., 111, 194, 199, 267-273, 276-312, 312-317, 458, 485-486, 494, 508, 520
 Pickens, H. D., 355, 396
 Pieters, C. E., 132, 193
 Pike, N., 146
 Pillet, R. A., 421, 472
 Pimsleur, P., 433, 435-436, 472
 Pinard, A., 296, 316
 Pincus, J., 240, 264
 Pintner, R., 11, 33-36, 102, 351, 353, 394
 Plana, 20
 Plank, E. N., 332, 394
 Plank, R., 332, 394
 Plato, 334
 Politzer, R. L., 414, 419, 457, 472
 Polya, G., 300, 316
 Postal, P., 414, 470
 Postlethwait, S. N., 229, 264
 Postman, L., 491, 520
 Powell, L. F., Jr., 583, 591, 595-600, 602-605, 607, 614-615, 621, 640
 Powers, F. F., 351, 353, 391
 Powers, S. R., 130-131, 181, 194
 Prator, C. H., Jr., 402, 472
 Pressey, L. C., 131-133, 150, 170, 174, 179, 192, 194; *see also* Cole, L.
 Pressey, S. L., 181, 203, 220, 222, 224, 264, 351, 356, 394
 Pribram, K., 273-274, 314
 Price, L. A., 56, 59, 98
 Prince, M., 321
 Pruzek, R. M., 69, 102
 Pryor, H. B., 349, 394
 Putnam, H., 82-83, 89, 102
 Putnam, J. J., 321-322
 Pyle, W. H., 351, 394
 Quensel, C. E., 532-533, 578
 Queresu, 66-67
 Quinn, T. J., 414, 417, 472
 Quirk, R., 423, 460, 473
 Rader, G., 130, 194
 Rajaratnam, N., 48, 97
 Ramseyer, J. A., 498, 505, 517
 Rand, E., 418, 473
 Rank, O., 374
 Rankin, P. T., 34, 102
 Rapaport, D., 322-324, 394
 Raph, J., 299, 316
 Rasch, G., 48, 102, 296, 316
 Rather, D., 82
 Ravem, R., 445, 473
 Ravitch, D., 82, 102
 Rayburn, W. G., 78, 102
 Redl, F., 366, 370, 394
 Reed, H. B., 131, 134, 164, 194
 Rehnquist, W. H., 583
 Reibel, D., 420, 471
 Reichert, D., 110, 115, 186
 Rejall, A. E., 138, 194
 Reller, T. L., 505, 520
 Reppucci, N. D., 240-241, 264
 Resnick, L. B., 85, 102, 229, 256, 264
 Reynolds, M. M., 349, 394
 Ribes-Inesta, E., 236, 262
 Rice, J. M., 125, 194
 Richards, J. C., 446, 473
 Richardson, J. R., 161, 193
 Richardson, M. W., 46, 100
 Richardson, S. C., 75, 102
 Richey, H. G., 160, 194
 Richman, B., 123, 185
 Ridenour, N., 358
 Rinsland, H. D., 114, 126, 128-129, 136-137, 141, 154, 181, 194
 Ripple, R. E., 276, 294, 316
 Rivenc, P., 411, 469
 Rivers, W. M., 417, 421, 424, 438, 441-444, 457, 473
 Rivlin, A., 614, 644
 Rivlin, H. N., 355, 369, 391, 394
 Robbins, I., 114, 198
 Roberts, D. A., 398, 473
 Roberts, L., 439, 472
 Robinson, E. S., 200, 264
 Robinson, F. P., 351, 356, 394
 Robinson, J. W., 170, 192
 Robison, R. K., 374, 394
 Rockcastle, V. N., 276, 294, 316
 Roden, A. H., 236, 262
 Rodriguez, D. P., 591, 601
 Roebor, E. C., 153, 194
 Roentgen, W. K., 386
 Roethlisberger, F. J., 499, 520
 Rogers, L., 302, 317

- Rosen, P. L., 606, 608, 644
 Rosenbaum, P. S., 419, 473
 Rosenthal, T. L., 247, 264
 Ross, D., 111, 194
 Roszkopf, M. F., 300, 316
 Rotter, J. B., 244, 264
 Roudebush, R. L., 355, 396
 Roulet, E., 416, 421, 473
 Rousseau, J. J., 350
 Rousselet, P., 613
 Rowe, C. E., 169, 194
 Royce, J. R., 85, 102
 Rubadeau, D. O., 299, 313
 Rubakin, N. A., 124
 Rubenstein, B. O., 365, 390
 Rubin, J., 432, 447, 473
 Rudy, W., 116, 194
 Rugg, H., 159
 Ruml, B., 11
 Ruskin, R. S., 225, 264
 Russell, D. H., 110, 121, 129, 151, 181, 188, 194
 Rutherford, W. E., 419, 473
 Ryan, W. C., 359, 366, 394
- Sabaroff, R. E., 116, 189
 Sacks, H., 422, 473
 Sacks, S., 586-587, 643-644
 Sandiford, P., 351, 353, 394
 Santos, E. H., 437, 469
 Sapon, S. M., 435-436, 467
 Saporta, S., 422, 473
 Saunders, D. R., 55, 102
 Saunders, J. T., 240-241, 264
 Sauvageat, A., 411, 469
 Savignon, S. J., 431, 473
 Scanlon, W., 618, 644
 Scarlet, W., 171, 193
 Scarr-Salapatek, S., 88, 102
 Schachtel, E. G., 494, 520
 Schachter, J., 446, 473
 Schaeffer, B., 246, 263
 Schauer, V. P., 175, 193
 Schegloff, E. A., 429, 473
 Scherer, G. A. C., 448-449, 453, 473
 Schilder, P., 330, 394
 Schmid, J., 55, 102
 Schoenfeld, W. N., 214, 225, 262
 Schoonover, T. L., 154, 190
 Schorling, R., 160, 176, 193
 Schramm, W., 157, 160, 193
 Schrupp, M. H., 363, 394
 Schultz, 67
 Schultz, T., 487
 Schumann, J. H., 445, 473
 Schwab, J. J., 398, 473
 Schwebel, M., 299, 316
 Schwesinger, G. C., 154, 193
 Scoggin, M., 171, 196
 Scott, F., 114, 193
 Scott, T. H., 494, 517
 Scovel, T., 441, 474
 Seago, D. W., 108, 115, 193
 Searl, M. N., 337, 394
 Searle, J., 428, 474
 Sears, R. R., 319, 324, 343, 394
 Seashore, H. G., 41, 65, 93
 Seashore, R. H., 110, 151, 154, 181, 193
 Sebeok, T. A., 406, 433, 472
- Seegers, J. C., 110, 131, 134, 161, 181, 193
 Segal, E. M., 491, 520
 Selinker, L., 415, 446, 474
 Serra, M. C., 130, 193
 Serrano, J. A., 597
 Serrano, J., Jr., 597
 Shaffer, L. F., 366, 368, 389, 394
 Shakow, D., 322-324, 394
 Shalala, D. E., 618, 644
 Shanner, W. M., 228, 261
 Shapiro, S. R., 174, 193
 Sharp, S. E., 6, 102
 Shaw, B., 376
 Shaw, J. C., 274, 315
 Shaycoft, M. F., 41, 98
 Shearson, G. F., 504, 518
 Sherman, J. G., 225, 262, 264, 483, 520
 Sherman, L. A., 167, 193
 Sherman, M., 366, 394
 Sherrington, C. S., 205
 Shoben, E. J., Jr., 355, 366, 368, 391, 394
 Sholty, M., 119, 193
 Showel, M., 170, 193
 Shulman, L. S., 293, 300, 316
 Shuy, R. W., 429, 474
 Sidman, M., 210, 264
 Sigel, I. E., 277, 316
 Silard, J., 607, 612-613
 Simm, A., 449, 474
 Simon, J., 7, 11, 15, 48, 72-75, 95, 102
 Simon, H. A., 256, 264, 274, 315, 501, 506, 520
 Simpson, R. G., 351, 394
 Sinclair, H., 278, 297, 302, 314, 316
 Sjöstrand, W., 543, 578
 Skinner, B. F., 199, 204-217, 220-230, 233, 242-245, 249, 260, 261, 264-265, 406, 443, 474, 483
 Skinner, C. E., 356, 394
 Skinner, M., 117, 193
 Skolnick, B. P., 453, 468
 Smith, B. O., 14, 59, 102, 512, 520
 Smith, K. J., 178, 193
 Smith, M., 100
 Smith, M. E., 113, 115, 193
 Smith, M. K., 113, 193
 Smith, N. B., 120, 160, 162, 168, 173-174, 181, 193
 Smith, P. D., Jr., 449, 453, 474
 Socrates, 224
 Solomon, C., 306, 315
 Sommer, R., 488, 520
 Sommers, A. A., 588, 644
 Sonquist, H., 305, 316
 Sorenson, 356
 Soule, G., 377, 394
 Spache, G., 181
 Sparks, J. N., 363, 394
 Spaulding, S., 181
 Spearman, C., 12-14, 21-24, 28, 38, 46, 51, 90, 102
 Spence, K. W., 204
 Sperling, M., 335, 394
 Spock, B., 380, 394
 Spolsky, B., 422-423, 431, 436, 474
 Stack, E. M., 410, 474
 Stagner, R., 169, 189
 Stanley, J. C., 46-48, 103
 Stanton, H. M., 351, 395
 Starch, D., 351, 353, 356, 395

- Start, K. B., 478, 520
 Stauffer, R. G., 121, 195
 Steere, G. H., 380, 395
 Steffe, L. P., 300, 316-317
 Stendler, C. B., 346, 349, 380, 393, 395
 Stenquist, J. L., 181
 Stern, C., 111, 195
 Stern, H. H., 397, 423, 439-440, 447, 449, 453, 471, 474
 Stern, W., 8, 16, 103
 Stern, W. L., 111, 195
 Sternberg, R. J., 59, 62, 103
 Stevens, A. L., 256, 263
 Stevens, S. S., 14, 60, 103
 Stevenson, P. R., 131-132, 195, 197
 Stevick, E. W., 438, 474
 Stewart, P., 583, 607, 616
 Stiles, L. J., 477-478, 520
 Stivers, P., 478, 520
 St. Lawrence, F., 174, 195
 Stockwell, R. P., 418, 436, 472, 474
 Stoddard, L. T., 210, 264
 Stoltz, L. M., 319
 Stolz, S. B., 230, 241, 265
 Stouffer, G. A. W., Jr., 362, 395
 Stout, R. T., 618, 644
 Strang, R., 171, 181, 196, 346, 348-349, 395
 Strayer, G. D., 585, 645
 Strevens, P., 407, 457, 469
 Strickland, C. E., 111, 150, 196
 Strickland, R., 181
 Strong, E. K., 351, 353-354, 395
 Stroud, J. B., 351, 353, 356, 395
 Sugarman, S. D., 589-590, 597, 643
 Sullivan, E. T., 28, 30, 103
 Sullivan, E. V., 247, 265
 Sullivan, H. B., 122, 134, 166, 187
 Sullivan, H. S., 322, 343
 Sullivan, M., 374, 395
 Sulman, A. M., 380, 395
 Sulzer, B., 236, 265
 Sundland, D., 436, 472
 Suppes, P., 60, 100, 103, 173, 186, 275, 317, 319, 450, 467, 480, 518
 Suzman, R. M., 319
 Svartvik, J., 423, 473
 Svensson, N. E., 71, 103, 529-530, 552-554, 579
 Swain, M., 441, 449, 474
 Sweet, H., 402, 458, 475
 Symonds, P. M., 129-130, 142, 154, 197, 364, 366, 373, 395, 403, 475
 Szymanska, A., 269, 316
 Taft, J., 374, 395
 Tanner, A. E., 346, 349, 395
 Tarback, S., 300, 316
 Tatsuoaka, M. M., 62, 103
 Taylor, C. W., 56, 103
 Taylor, D. M., 437, 469
 Taylor, F. W., 498, 520
 Taylor, W. L., 6, 103, 179, 196
 Temp, G., 170, 196
 Terman, L. M., 7, 11, 15, 22, 30, 103, 110, 115, 196, 296, 317, 351, 481-482, 508-509, 517, 520
 Terrace, H. S., 210, 265
 Tharp, J. B., 136, 169, 181, 196
 Tharpe, R. G., 238, 265
 Thomas, D. S., 380, 395
 Thomas J. A., 588, 645
 Thomas, W. I., 380, 395
 Thompson, G. G., 346, 349, 395
 Thompson, W. R., 494, 519
 Thomson, G., 10, 73, 95
 Thoresen, C. E., 238, 242-244, 247, 249-250, 262, 265
 Thorndike, E. L., 8, 11, 17, 22, 24, 36, 46, 103, 107-110, 113-115, 118-131, 134, 138-157, 160-167, 170, 172, 177-178, 180-183, 187, 197, 199-204, 212, 224, 265, 320, 349, 351, 381, 383-385, 395, 403, 448-449, 458, 475, 483-485, 489, 501, 508, 520-521
 Thorndike, R. L., 40-41, 89, 103, 155, 197
 Thorpe, L. P., 170, 197, 366, 369, 395
 Thralls, Z. A., 175, 197
 Thurstone, L. L., 8, 13, 22-24, 31-32, 36-40, 44, 48, 51-52, 55-57, 66, 86-87, 89, 103, 322, 395
 Thurstone, T. G., 31-32, 36, 40, 104
 Tidyman, W. F., 181
 Tiegs, E. W., 28, 30, '03, 356, 395
 Tireman, L. S., 139, 197
 Titchener, E. B., 6
 Todesco, A., 447, 471
 Tolman, E. C., 204
 Tonks, H. L., 351, 395
 Toops, H. A., 24, 35, 104
 Trager, E. C., 419, 465
 Trager, G. L., 404, 465
 Travers, R. M. W., 492, 521
 Traxler, A. E., 174, 197
 Trim, J. L. M., 428, 475
 Troen, S. K., 117, 197
 Trow, M., 355
 Tucker, G. R., 449, 453, 470
 Tucker, L. R., 44, 47, 55, 104
 Tuddenham, R. D., 296, 317
 Tuinman, J. J., 121, 197
 Tunstall, K., 436, 470
 Tversky, A., 60, 100
 Twaddell, W. F., 197; *see also* Twaddell, W. F.
 Twaddell, W. F., 140, 169; *see also* Twaddell, W. F.
 Tyack, D. B., 116-117, 197
 Tyler, J. M., 349, 395
 Tyler, L. I., 331, 396
 Tyler, R. W., 64-65, 85, 104
 Uhl, W. L., 143, 146, 157, 162-163, 166-167, 181, 197
 Ullmann, L. P., 204, 229, 262, 265
 Undeutsch, U., 550, 579
 Updegraff, H., 585, 645
 Upton, L., 175, 197
 Urwick, L., 499, 519
 Usdan, M. D., 618, 643
 Valdman, A., 457, 475
 Valentine, C. W., 351, 353, 396
 van Buren, P., 418, 426, 475
 Vandenberg, S. G., 88, 104
 Vander Beke, C. E., 148, 181, 197
 van Ek, J. A., 428, 475
 Van Wagenen, M. J., 181
 Varadi, T., 446, 475
 Varma, V. P., 299, 317
 Viereck, G. S., 376, 396
 Vincent, E. L., 346, 349, 389
 Vincent, P., 618, 644

- Vogel, M., 171, 179, 197
 Voyat, G., 307, 313
 Wachs, H., 302-303, 313
 Waddell, C. W., 346, 349-350, 396
 Waddle, C. D., 419, 465
 Wade, N., 82, 104
 Walker, D. A., 17-18, 49, 104
 Walker, H. M., 19-20, 105
 Wallace, J. G., 277, 317
 Wallace, W. L., 44, 70
 Wallen, N. E., 492, 521
 Wallin, J. E. W., 365-366, 396
 Wang, C. K. A., 154, 197
 Wang, M. D., 47, 103
 Ward, C. F., 148, 197
 Ward, J. L., 131, 197
 Wardhaugh, R., 414-415, 475
 Wardwell, F., 168, 192
 Ware, A. E., 175, 197
 Warren, H., 319
 Warren, R., 24, 105
 Washburne, C. W., 134, 159-160, 171, 179, 181, 193, 197, 228, 265
 Watson, G., 355, 357, 391, 396
 Watson, J. B., 92, 205, 323, 349, 354, 501
 Wattenberg, W. W., 366, 370, 394
 Watters, G., 158, 198
 Weeks, H. F., 355, 396
 Weikart, D., 302, 305, 317
 Weinfeld, F. D., 588, 643
 Weinreb, J., 364, 396
 Wells, 8
 Werdelin, I., 56, 105
 Wertheimer, M. A., 448-449, 453, 473
 Werts, C. E., 62, 105
 Wesche, M. B., 397, 447, 475
 Wesley, E. B., 162, 193
 Wesman, A. G., 41, 65, 95
 West, L. J., 138-139, 169-170, 187, 198
 Wetzel, R. J., 238, 265
 Wherry, R. J., 49, 61, 105
 Whipple, G., 159, 181, 198
 Whipple, G. M., 7, 103, 127, 163, 181, 198, 228, 265
 Whitbeck, R. H., 175-176, 198
 White, 87
 White, B. R., 583, 596, 598, 604, 607
 White, M. N., 126, 185
 White, R., 595, 640, 643
 White, R. K., 482, 492, 500, 519
 White, R. W., 493-494, 521
 White, W. A., 321-322, 358, 360, 366, 470, 396
 Whitebrook, J. S., 247, 264
 Whitley, M. T., 346, 349, 393
 Wickman, E. K., 358, 362-363, 394
 Widdowson, H. G., 428-429, 465, 475
 Wienckowski, L. A., 230, 265
 Wilder, D. E., 147, 184
 Wiley, D. E., 82, 99, 588, 645
 Wiley, W., 164, 198
 Wilkerson, D. A., 78, 98
 Wilkins, D. A., 415, 419, 421-424, 426, 428, 475
 Williams, M. F., 618, 644
 Williams, P., 299, 317
 Willoughby, L., 227, 261
 Wilson, E. M., 175, 198
 Winer, B. J., 49, 105
 Wirth, A. G., 116, 198
 Wise, A. E., 478-479, 521, 588-590, 605, 607, 614-615, 644-645
 Wise, C. T., 135, 142, 198
 Wishy, B., 111, 198
 Wissler, C., 6, 105
 Witherington, 356
 Wittels, F., 374, 396
 Witty, P., 181
 Woerdehoff, F. J., 449, 467
 Wohlfarth, J. H., 197
 Wolf, B. L., 588, 644
 Wolf, E. P., 608, 614, 645
 Wolf, M. W., 233, 260
 Wolf, R. M., 85, 104
 Woffle, D., 50-51, 105, 527, 579
 Wolinsky, S., 613-614
 Wood, B. D., 35, 50, 95, 105, 181
 Woodrow, 11
 Woodruff, 356
 Woodworth, R. S., 323, 483-485, 508, 520-521
 Woody, C., 181, 351, 396
 Woodyard, E., 10, 103
 Worcester, D. A., 355, 396
 Wright, S., 61, 105
 Wrightstone, J. W., 114, 198
 Write, S., 609
 Wundt, W., 5-6, 15, 105
 Yalden, J., 457, 475
 Yang, 507, 509
 Yang, W. C., 618, 644
 Yates, A., 543, 579
 Yauch, W. A., 500, 521
 Yerkes, R. M., 12, 22, 42, 61, 68, 105
 Yoakam, G. A., 129, 134, 136, 160, 166, 168, 176-177, 181, 198
 York, R. L., 588, 643
 Young, J. W., 177, 198
 Young, M., 79, 81, 105
 Yourman, J., 362, 396
 Yule, G. U., 19, 61, 105
 Zachry, C. B., 331, 396
 Zimmerman, B. J., 247, 264
 Zimmerman, W. S., 41, 70, 99, 154, 189
 Zink, N. E., 162, 175-176, 193, 198
 Zinnes, J. L., 60, 103

SUBJECT INDEX

- 11+ examination (England), 79, 542, 545, 550
 1940 School Commission (Sweden), 525-526, 532, 544-548, 551-552, 574
 1946 School Commission (Sweden), 526-527, 545, 548
 1955 University Commission (Sweden), 539, 541
 1957 School Commission (Sweden), 528-530, 548-550, 552, 554, 575
 study of school curriculum, 555-571
 1960 Gymnasium Commission (Sweden), 571-574
- Ability grouping, 71-72, 80, 543
 Ability reserve, 525; *see also* Talent reserve.
 Absolute scaling methods (Thurstone), 23
 Academic aptitude, 115
 Achievement tests, 64, 67, 76, 247-249
 Admissions testing program for colleges, 35-36, 38, 42, 82
 Adult education
 and second-language education, 448-449
 Thorndike's contributions to, 180
 Adult vocabulary, 109-111, 113
 Aggression, 336, 383
 Algebra, educational research and, 177-178
 Alpha factor analysis (Kaiser and Gaffrey), 55
 American Association for Adult Education, 180
 American Association of Psychiatric Clinics for Children, 343
 American Association of School Administrators (AASA), 477
 American College Testing Program (ACT), 44, 70, 77
 American Council on Education (ACE), 31
American Council on Education Psychological Examination, 25, 31-32
American Council on Education Psychological Examination for College Freshmen, 8
 American Council of Learned Societies (ACLS), 404
 American Council on the Teaching of Foreign Languages (ACTFL), 455-456
 American Heritage Intermediate Corpus, 141
 American Heritage School Dictionary, 123
 American method (teaching with textbooks), 157-158
 American Orthopsychiatric Association, 343
 American Psychoanalytic Association, 322
 Analysis of variance, 48
 Analytical methods of rotation, 55
 Anna Freud's *Lectures to Teachers*, 325-327
 Annual vocabulary, 109
 Anomaly for a paradigm (Kuhn), 281
 ANOVA (analysis of variance), 48
 Anthropometric Laboratory (Galton), 6, 21
 Anti-intellectualism, 508
 Application model, 460
 Applied linguistics, 405, 408-409, 454-455
 Army Air Force Qualifying Test, 40
 Army Alpha Examination, 8-9, 12, 35, 43, 68
 Army Beta Examination, 8-9
 Army General Classification Test (AGCT), 40
 Army Language School, 405
 Army Method language courses, 405
 Army testing program, 22
 Assimilative reading, 132
 Association Internationale de Linguistique Appliquée (AILA), 455
 Association
 Law of, 402-403
 statistical concept of, 20
 Associationism, 200, 212
 conception of learning, 489-490
 Attentional processes, 253-254
 Attenuation paradox, 47
 Attribution theory, 240
 Audiolingualism, 407, 418-419, 421, 442, 452
 approach, 450
 method, 405, 414
 tests, 412
 Auxiliary hypotheses (Lakatos), 282-288
 Aversive stimulus, 207
 Avoidance, 237
- Baby box, 215
Baby and Child Care (Spock), psychoanalytic influence in, 380-381
Baker v. Carr, 589
 Baseline measure, 231
 Basic English, 139
 Basic research
 accountability in, 508-509
 on cognitive development, 485-486
 definition of, 480
 effect on attitudes toward giftedness, 480-482
 effect on classroom environment, 482-483
 effect on classroom leadership style, 482
 effect on classroom practices, 480-487
 effect on curriculum, 483-486
 effect on paradigms of human behavior, 487-497
 effect on professional preparation in education, 497
 identification of problems in, 507-508
 Skinner's contributions to, 483
 and the university, 506-509
 Basic research in education, 477-517
 support for, 478
 Behavior, 234, 415
 control, 232
 theory, 255-256
 Behavior modification, 229-244, 251, 256; *see also* Behaviorism.
 and the autistic child, 238
 counterinfluences, 244-257
 in educational settings, 235-239
 extrapolation of animal experiments to human behavior, 214-215
 general approach, 230-233
 government report on, 241
 important publications in, 229-230, 233, 236
 in natural settings, 233-235, 240
 and the retarded child, 238
 techniques, 230-231, 235, 240-241
 in token economies, 233-235

- Behavioral analysis of language learning, 435
 Behavioral models, 247
 Behavioral objectives, 256
 Behavioral programming, 243
 Behaviorism, 414; *see also* Behavior modification.
 in language learning, 422
 and language teaching, 433-434
 theory, 257, 413
 Behaviorist learning theory, 412-413
 Behavioristic theory of L2 learning, Chomsky's
 criticism of, 406
 Bilingualism, 406, 445, 447
 in Canada, 432-433
 Binary scale (pass-fail), 15-16
 Binet's mental age scale, 14-15, 27
 Binet-Simon scale, 7
 Binet-type items, 16
 Binet-type scales, 7, 15, 27, 48
 Binet-type tests, 92
 Biofeedback, 229
 Bivariate descriptive statistics, 19-20
 Brandeis brief, 608
 British associationism, 200
 British grammar school, 531
 British Intelligence Scale, 296
Brown v. Board of Education, 589
*Bulletin of the Institute for Research in English
 Teaching (BIRET)*, 139
Buros Yearbooks, 27, 38, 40, 64, 66, 68-69, 153

California Tests of Mental Maturity, 25, 28-31
 Canonical correlation, 61
Care and Feeding of Children (Holt), 380-381
 Carroll-Sapon Modern Language Aptitude Test
 (MLAT), 435
 Case grammar, 427
 Cathexis, 330
 Cattell's hierarchical model of intellectual abilities,
 58, 85
 CAVD test (Thorndike), 9, 154
 CEEB, 43-44; *see also* College Entrance
 Examination Board.
 Center for Applied Linguistics (CAL), 407, 413,
 454-455
 Center for Genetic Epistemology, 272
 Center for Research on Language and Language
 Behavior, 455
 Centre de Recherche et d'Etude pour la Diffusion
 du Français (CREDIF), 456
 Centre for Information on Language Teaching
 (CILT), 455, 457
 Centroid method, 51, 53
 Child-centered school, 371-372
 Child-development movement, 340, 342-350
 history of, 342-343
 textbooks, 343-350
 Child guidance clinic movement, 343
 Child Study Association of America, 380
 Child-study movement, 111, 340, 342-350
 history of, 342-343
 textbooks, 343-350
 Childhood sexuality, 385
 Children's vocabularies, 113, 127-129
 Chimpanzees, teaching language system to, 255
 Circular classroom, 488, 492-493
 Clark's self-concept studies, 487
 Classical test theory, 46-47
 Classroom architecture, 488-497
 discrepancy between design and use, 496-497
 in relation to conceptions of learner, 488-497
 Closure, 493, 591
 Cloze test, 6, 179
 Coefficient alpha (Cronbach), 47, 49
 Cognitive factors, 56-59
 Cognitive information processing, 255
 Cognitive latent learning, 213
 Cognitive processes, affective factors, 33, 276
 Cognitive psychology, 13, 58, 224, 240, 244, 250,
 253-257, 407, 420
 and educational problems, 255-257
 and language teaching, 255-256
 Cognitive theories, 213, 244-247, 420
 of performance, 257
 Coleman Report, 610
 College admissions, 35-36, 38, 42, 82, 115
 in United States, 79-80
 College Board Scholastic Aptitude Test (SAT), 38,
 44, 67-71
 effects of coaching, 78
 College Entrance Examination Board (CEEB), 9,
 35-36, 38, 42, 44, 67-71, 77, 82
 College entrance tests, 35-36, 38, 42, 115
 Commission on Student Welfare (Sweden), 532
 Committee on Curriculum and Examination in
 Secondary Schools, 74-75; *see also* Norwood
 Report.
 Committee on the Language Program (CLP), 405
 Committee of the Textbook of the National
 Society for the Study of Education, 160
 Common factor, 51-52
 variance, 51
 Common-factor space, 54
 Communalities
 of tests, 51
 of variables, 54-55
 Communication networks, 492
 Communicative competence, 424, 427-428,
 430-432
 Comparison philosophy in school policymaking in
 Sweden, 547-548, 554-555
 Compensatory education programs, 301
 Competitive school system, 527-528
 Completion test, 6
 Componential analysis, 426
 of mental tasks (Sternberg), 59
 Comprehension, 120, 163
Comprehensive Abilities Battery, 42
 Comprehensive school, 76
 objectives and content of curricula in Sweden,
 555-557
 secondary level, 75-76
 "streaming" in England, 543
 in Sweden, 541, 543, 545-549, 574-576
 Compulsory schooling (Sweden), structure of,
 541-543
 Computational technology, 24, 61
 Computer-assisted instruction, 229
 Concrete operations stage, 270
 Concurrent validity, 64
 Conditioned reflexes, 330
 Conference on Cognitive Studies and Curriculum
 Development, 276
 Configuration, 591
 Confirmatory factor analysis (Joreskog and
 Lawley), 55
 Conflict, role in developmental process, 327

- Connecticut Society for Mental Hygiene, 358
 Conservation properties of objects (Piaget), 485-486
 Constancy of intelligence, 79
 Constitutional traits, 58
 Constructive analysis, 294, 301-302, 306
 Context, 121, 146
 Contingencies of reinforcement, 217, 236
 Contrastive analysis, 406, 418, 429, 446, 460
 Contrastive linguistics, 460
 Convergent model, 460
 Cooperative English Tests, 154
 Cooperative Research Act (1954), 477
 Cooperative Research Program, 478
 Cooperative Test Service, 154
 Core-Vocabulary Readers, 162
 Correction for attenuation, 21
 Correlated factors, 51
 Correlation, 20
 coefficient, 20, 51
 Galton's notion of, 6
 method of, 61
 multiple, 61
 theory of, 21
 worksheets for computing, 24
 Covert self-control, 249-250
 and test anxiety, 250
 Covert self-instruction, 250
 Covert self-reinforcement, 249-251
 Cronbach's coefficient alpha, 49
 Cross-lagged panel correlation, 62
 Crystallized intelligence (Cattell), 58, 85
 Curriculum
 general-education requirements, 571-574
 psychoanalytic view, 331
 Curriculum reform, Sputnik and, 275
 Curriculum studies in Sweden, 555-574

 Dale-Chall readability formula, 153
 Darwinism, 5, 111
 Debugging, 309
 Decalages, 287, 296
 Deep-structure cases, 427
 Defense mechanisms, 332, 351, 354, 368-370, 385
 Democratic education, 5, 115-122
 in Sweden, 527
 Derived measurements, 59-60
 Descriptive linguistics, 407
 impact on second-language education, 460
 primacy of speech, 410-411
 Developmental approach, 11
 Developmental psychology, 252
 Dewey
 attitude toward psychoanalysis, 382-383
 Laboratory School, 490
 linking science of, 201-202
 views on elementary education, 381-383
 views on relationship between instincts and habits, 381-383
 Diagnostic profile, 28
 Dictionaries, 122-124
 Differential Aptitude Tests (DAT), 41, 65-67
 Differential reinforcement, 208-209, 245
 Differentiation
 ability grouping, 543
 class, 543
 homogeneous grouping, 543
 of responses, 208-209
 by school type, 542
 sectioning, 543
 streaming (England), 543
 track, 542-543, 549
 Differentiation in schooling in Sweden, 541-555
 arguments for early and late transfer to *realskola*, 544-545
 in educational reform, 543-549
 elective system, 548
 organizational, 542-543, 545, 548-549, 552-553
 pedagogical, 542-543
 role of individual differences, 542-543
 screening, 550
 selective system, 548
 and social background, 549-555
 Discipline
 and instinctual impulses, 335
 psychoanalytic view of, 334-336
 Discourse analysis, 424, 429
 Discovery learning, 293, 300
 Discrimination, 209-210, 213, 237
 Discrimination learning, 209
 Discriminative stimulus, 209, 243
 DISTAR program, 229
 Distractors, 114, 154-155
 District Power Equalizing (DPE) in school finance, 590-591
 Doctrine of formal discipline, 483
 Dolch list, 149
 Douglas v. California, 589
 Drive- or tension-reduction theory of behavior, 493
 Dropouts (Sweden), 550-555, 575
 relationship between social class and, 552
 Dynamo (Faraday), 508

 Economics of education (Schultz), 487
 Education
 Anna Freud's views on, 325-327
 association with psychology, 201-204
 basic research in, 477-517
 conflict between child's instincts and aim of, 325-327
 effects of forceful imposition on child's development, 326-327
 evidence of impact of psychoanalysis, 340-381
 as form of repression, 390
 indulgence and deprivation as means of, 335-336
 limits of, 336, 388
 as a practical science, 202-204
 psychoanalytic theory relevant to, 324-340
 psychoanalytic view of discipline in, 334-336
 psychoanalytic view of limited goals, 327
 psychoanalytic view of moral development and, 334-336
 relationship to psychology, 199-204
 role of fear of losing love in, 328-329
 role of internal structures in, 335
 role of reward and punishment in, 328-329
 theory-oriented research in, 477-517
 Education Commission of the States, 617
 Education Index, 114-115, 139
 Educational administration
 and fusion process, 502
 idiographic (individual) style, 503
 nomothetic-idiographic model, 504-506
 nomothetic (normative) style, 503

- shift from prescriptive solutions to theoretical inquiries, 503-506
- shift from solution orientation to inquiry orientation, 503
- shift from vocational principles to theoretical concepts, 501-503
- social system model, 504-506
- transactional style, 503
- Educational administrators
 - effects of conceptual shifts on textbooks used in preparation of, 504-506
 - and human relations management, 499-501
 - impact of research on preparation of, 498
 - and scientific management, 498-499
- Educational linguistics, 454
- Educational measurement, 22
- Educational opportunities (Sweden), disparities in, 541-555
- Educational policy
 - judicialization of conflict, 581
 - and social science, 581-620
- Educational policymaking, 523-531
 - developmental psychology and, 574
 - social science as instrument of advocacy in, 581-620
- Educational practice, Piagetian theory and, 267-312
- Educational psychology, 202-204, 384
 - textbooks, 203-204
- Educational psychology movement, 340-341, 350-358
 - history of, 350-351
 - student and teacher attitudes, 357-358
 - textbooks, 351-357
- Educational psychology (Thorndike), 384
- Educational reform in Sweden, 523-577
- Educational research
 - consequences of avoiding theory, 511-513
 - contribution to decision making, 576
 - educational reform and, 523-577
 - effects on educational policy, 487-497
 - examples in second-language education, 448-451
 - funding, 530
 - genetic bias, 510
 - impact on second-language education, 399-401, 462
 - legitimation as function of, 182-183
 - objections to utilizing theory, 513-514
 - organism-centered bias, 510
 - planning function in policymaking, 576
 - purposes of, 530-531
 - relationship to practice, 397-401
 - sociological bias, 510
 - theoretical bias, 510-511
 - theoretical safeguards, 514-515
 - views concerning its focus, 509-515
- Educational research movement, 117
- Educational Testing Service, 43-44, 67, 77, 82, 403
- Ego function, 331-332, 335
- Ego psychology, 324-325, 334, 387
- Elementary education, psychoanalysis and, 319-388
- Elementary English Review, 122
- Elementary Modern Language Aptitude Test (EMLAT), 435-436
- Elementary and Secondary Education Act (1965), 477
- Empirical law of effect, 212
- Empiricus, 172
- Employment selection tests, legislation and regulation, 80
- Energy (psychic), 329-330, 332
- English Language Institute, 406, 456
- English as a second language (ESL), 138
 - program, 406
 - teaching, 138-139, 141
- English Teaching Information Centre (ETIC), 456
- Environmental planning, 243
- Environmental space, 258
- Equal-protection analysis, 596, 599, 602-604, 614, 616-617
 - use in school-finance reform, 582, 588-589, 591-592
- Equality of educational opportunity, 527
 - for advanced education, 531-541
 - in Texas, 582-583
- Equivalent-form reliabilities of tests, 26, 64
- ERIC Clearinghouse on Languages and Linguistics, 457
- Errorless learning, 212
- Esperanto, 139
- Ethnography of communication, 430
- Études d'Épistémologie Genetique, 272, 278
- Evolution
 - Darwinian theory of, 5, 111
 - Lamarckian theory of, 328
- Experimental design, 62
- Experimental psychology, 203
- Experimental sections in tests, 42
- Exploratory factor analysis (Jöreskog and Lawley), 55
- Extinction, 207, 210, 234, 237
- Factor analysis, 19, 21-23, 46, 50-56, 85-88, 90, 92-93, 435
 - alpha, 55
 - British school of, 55
 - confirmatory, 55
 - exploratory, 55
 - image, 55
 - linear model of, 52
 - methodology of, 54-56
 - nonlinear model of, 52, 55
 - studies in cognitive abilities, 56
 - three-mode model, 55
 - use of computers in, 52, 56
- Factor domains, 56
- Factor model, 53
- Factors, proliferation of, 57
- Fading, 210, 232
- Fear reduction, 237
- Ferguson's learning-theoretic interpretation of factors, 58
- Field Initiated Studies Program, 478
- Field theory (Lewin), 492-493
- Figurative structures, 301-302
- Fiscal neutrality, 595
 - in school-finance reform, 590
- Flanagan Aptitude Classification Tests (FACT), 41
- Fleckeisen's word-frequency principle, 124
- Fluid intelligence (Cattell), 58, 85

- Folkshola*, 561
 Ford Foundation, 453-454, 617
 Foreign language
 aptitude, 435
 curriculum, 149
 teaching, 124, 140, 142, 149, 168
 Foreign Language in the Elementary School (FLES), 439-440, 453
 Forensic social science, 614
 Formal discipline
 doctrine of, 483-485
 Thorndike-Woodworth studies, 483-485
 Formal operational level (Piaget), 485-486
 Formal-operations stage, 270
 Foundation plan of school finance (Strayer and Haig), 585, 593-594
 Free-association study of writing vocabularies, 135
 Free problem-solving activities, 293, 306
 Free-textbook laws, 158-159
 Frequency research, 124, 134
 Freud
 comparison with Darwin, 375
 comparison with Dewey and Thorndike, 381-385
 comparison with Einstein, 376-377
 comparison with Marx, 377
 impact on American culture, 374-381
 impact on art, 378
 as literary figure, 377
 Freud's (Anna) *Lectures to Teachers*, 325-327
 Fulbright exchange program, 454
 Full state assumption of school finance (Morrison), 585-586
 Functional analysis, 428-429
 Fundamental measurement (Campbell), 59
 Fundamental right, 598-600
 in school-finance litigation, 591-592
 Gaussian law of error, 19; *see also* Normal probability distribution
 General classification test, 40
 General-education requirements (Sweden), 571-574
 General factor in intelligence (g), 12-13, 30, 51, 86-87
 General lexicon, 109
 General word knowledge, 115
 Generalizability theory, 48
 Generalization, 213, 239
 Generative grammar, 414-421
 definition of, 418-419
 Generative semantics, 414-415
 Genius, Galton's studies of, 6, 9
 Geographical variability, 147
 German oral teaching method, 157-158
 Gestalt, 491
 Giftedness
 attitudes toward, 480-482
 effect of basic research on attitudes toward, 480-482
 Terman's longitudinal study, 481-482
 Gloss, 122
 Good Companion Series, 152
 Gooficon, 443
 Grade-repeating (Sweden), 550-555, 574
 Grammatical competence, 430
 Graphical methods of rotation, 55
 Greek, role in curriculum, 484-485
 Griffin v. Illinois, 589
 Group dynamics, 492, 500-501
 Group Examination Beta, 8
 Group factors in intelligence, 12, 23, 51
 Group intelligence tests, 3, 7, 12, 16-17, 25, 28, 37, 53, 62
 Group processes, 492
 Grundskolan, 555
 Guilford's Structure of Intellect (SI) Model (taxonomic model), 41, 57-58, 85, 486-487;
 see also Structure of Intellect Model
 Guilford-Zimmerman Aptitude Survey, 41, 154
 GUME Project, 449
 Guttman scale, 18
 Gymnasium, 528, 536, 538-539, 544-546, 552, 558-559, 561, 563-564, 567, 573
 curriculum, 571-574
 in West Germany, 550
 Hawaii, school finance in, 582, 586
 Hawthorne studies, effect on educational administrators, 499-500
 Hebb's sensory deprivation studies, 487
 Henmon-Nelson Tests of Mental Ability, 25-28, 62
 Heredity vs. environment, 10, 58, 72, 82, 87-89, 91, 94
 Heterogeneous test, 18, 49-50
 Hierarchical factor structure, 55
 Hierarchical ordering of behaviors, 238
 Hig (Walker), 17
 High-level aptitudes project (Guilford), 56
 Hobson et al. v. Hansen et al., 609-610
 Homeostatic model of self-maintenance, 493
 Homogeneous grouping of students, 34, 526, 543, 611
 Homogeneous test, 18, 49-50
 Hort (Vienna), 325
 Horton v. Meskill, 604, 616-617
 Human engineering, 253
 Human factors research, 252-257
 Human Nature and Conduct (Dewey), 381-383
 Human Nature and the Social Order (Thorndike), 384-385
 Human performance, 253, 257
 Human relations management, 503
 applications to educational administration, 499-501
 Hunt's information processing model, 59
 Hunt-Uzgrigis scale of infant development, 296
 Image factor analysis (Guttman), 55
 Imagery, 246, 250
 and learning, 250
 and memorization, 250
 Immersion programs, 450
 Incompatible alternative, 237
 Indeterminacy, 54-55
 Indiana University Linguistics Conference (1953), 406
 Individual differences, 53-54, 59, 276-277, 385, 434
 in language-learning processes, 446-447
 in mental ability, 5
 typologies, 14
 Individualization of instruction, 222, 228-229, 251, 611

- pedagogical differentiation in Sweden, 543, 546
- Industrial Council for Social and Economic Studies in Sweden (SNS), 528
 - curriculum study, 555-556
- Industrial Revolution, 5
- Infantile amnesia, 326
- Infantile sexuality, 345, 380
- Inferential statistics, 19
- Information processing, 10, 58-59, 252-257
 - model, 59
 - theories, 247, 253
- Inhibition, 331-332
 - intellectual, 332-333
- Insight, 491
- Institute of Educational Research (Stockholm), 528, 556
- Institute of Research in English Teaching, 139
- Instructional psychology, 257
- Intellectual ability, 1-105; *see also* Mental ability.
- Intellectual inhibition, 332
- Intelligence, 2, 9-10, 16, 38, 53, 91-94, 115, 119, 276-277, 458
 - constancy of, 79
 - definition of, 10-12
 - Dewey's views, 382
 - genetic basis, 81-84
 - group tests of, 3, 7, 12, 16, 25, 62
 - metric scale of, 15
 - sampling theory of, 10
 - two-factor theory of (Spearman), 12
- Intelligence quotient (IQ), 2, 8, 15-16, 38, 46
 - concept introduced by Stern, 8
- Intelligence tests, 3, 37, 115
 - impact on American educational system, 76-77
 - impact on British educational system, 72-76
 - in school selection process in Sweden, 553-554
 - use in school, 71-79
- Intelligence tests vs. achievement tests, 76-77
- Intensive Language Program, 404
- Interlanguage, 445-446, 448
- Intermittent reinforcement, 237
- International Center for Research on Bilingualism, 455
- International language, 139
- International Phonetic Alphabet (IPA), 402
- International Primary Factors Battery*, 42
- Interpersonal cohesion, 492
- Interval scales, 60
- Interval schedule, 207
- Intrapsychic approach, 230
- Iowa Child Welfare Research Station, 343
- Iowa Studies, 116
- "IQ Myth" (CBS television program), 82
- Item
 - difficulty, 47, 49
 - homogeneity, 17
 - selection, 63
- Item analysis, 22, 42-43, 46
- Item-characteristic function, 48-49
- Jargon, 512
- Jensen Report, 81-82
 - Kamin's views on, 87-88
- Job analysis, 89
- Judicial intervention, 596
- Keating Report, 462
- Keller Plan, 225-228
- Kibbutz education, 320
- Kindergarten List, 127
- Kohlberg's stage theory, 298-299
- Kuder-Richardson reliability coefficients, 49
- Kuhnian paradigms, 280-282
- L (Linguistic) scores, 32
- Laboratory School (University of Chicago), 201, 371
- Lamarckian theory of evolution, 328
- Language, 107-198
 - acquisition of, 113, 255-256
 - American Indian, 404
 - centers, 455
 - evolution of, 111
 - formation rules of, 430
 - pathology, 447
 - pedagogy, 425
 - policy, 432
 - speaking rules of, 430
 - testing, 403
- Language acquisition device (LAD), 442
- Language arts, 136
- Language departments in universities, 462
- Language Development Program, 452
- Language laboratory, 450
- Language learning, 256, 421-422, 435
 - behavioristic interpretation, 422
 - role of creativity in, 415
- Language Study Center of the Philippine Normal College, 454
- Language teaching
 - direct method, 402-403
 - effect of World War II on, 404-405
 - historical influence of linguistics in, 401-402, 404-407
 - historical influence of psychology in, 401-403, 405-407
 - notional approach to, 428
- Latency period in child development, *anc.*
 - learning, 326-327
- Latent trait, 48, 51
- Latin
 - role in curriculum, 484-485
 - word list, 123-124
- Law of Association (Sweet), 402-403
- Law of Effect (Thorndike), 201, 212-213, 384
- Law of Exercise (Thorndike), 384
- Law of Readiness (Thorndike), 384
- Laws of learning, 201
- Le Français Fondamental, 410
- Leadership style study, 482-483
- Leadership styles in classrooms, 482
- Learner
 - as active organism, in square classroom, 491-492
 - associationist view of, 489-490
 - conceptions of, 488-497
 - as empty organism, in rectangular classroom, 489-490
 - field-theory view of, 492-493
 - Gestalt view of, 491-492
 - and open classroom, 495-497
 - personalistic view of, 491-492
 - as problem-finding organism, 495-497
 - as social organism, in circular classroom, 492-493
 - social-psychological view of, 492-493

- as stimulus-seeking organism, in open classroom, 495-497
- stimulus-seeking paradigm, 495-497
- underlying paradigm, 493
- Learning**
 - competences, 58
 - drive-reduction paradigm, 493-494
 - group-centered, 492-493
 - inquiry-centered, 495-497
 - laws of (Thorndike), 330
 - learner-centered, 491-492
 - and performance, 246
 - preconditions for, 328
 - psychoanalytical concepts in, 329
 - readiness, 328
 - as reduction of drives, 493
 - role of fantasies in, 333-334
 - role of internal structures in, 328-329
 - role of motivation in, 329-330
 - as seeking of stimulation, 493-494
- Learning theory**, 256, 354, 384, 434, 444
 - in language training, 405
 - psychoanalytic view, 327-331
 - Thorndike's, 200-201
- Legitimation**, 182
- Lewin's group climate studies**, 492
- Lewin's leadership studies**, 492, 500-501
- Lexical competence**, 425
- Lexicon**
 - general, 109
 - subcultural, 147
 - technical, 109
- Libidinal fixations**, 329
- Linguistic analysis**, 412-413
- Linguistic competence**, 424
- Linguistic Society of America**, 404, 455
- Linguistic theory**
 - goals of, 417-418
- Linguistic universals**, 417-418
- Linguistics**, 410-433
 - anthropological, 404-405
 - applied, 408-409
 - contrastive analysis, 406, 413, 418, 446
 - discourse analysis, 429
 - formal analysis, 411-412
 - impact on language tests, 413-414
 - relationship between language teaching and, 421-423
 - sociolinguistic analysis, 430-433
 - structural, 404-405
- Linking science (Dewey)**, 201-202
- Listening comprehension skills**, 444
- Lock-step pedagogy**, 182
- LOGO environment**, 306-311
 - debugging, 309
- Large-Thorndike Semantic Count of English Words**, 151
- Lycée (France)**, 550
- Mass education**, 116
- Mastery learning**, 229
- Mathematics learning, emotional factors in**, 332-333
- Max-Planck-Institut für Bildungsforschung (Berlin)**, 531
- McGuffey Readers**, 120-121
- McInnis v. Shapiro**, 589, 617
- Meaning load**, 152
- Meaningful reading**, 121-122
- Measurement; see also Educational measurement and Psychological measurement**
 - computers in, 61
 - fundamental (Campbell), 59
 - psychological, 60
 - Stevens' definition of, 60
 - theory of, 3, 14-18, 59
- Measurement Research Center, University of Iowa**, 63
- Measurement scales**, 60
 - interval, 60
 - nominal, 60
 - ordinal, 60
 - ratio, 60
- Medical model**, 230, 240
- Memory span test**, 11
- Mental ability**, 1-105; *see also* Intelligence
 - influence of research on educational policies, 71-76
 - notions of, 9-14
 - theories of, 44-46, 85-87
- Mental ability test scores**, 86-87
- Mental ability tests**, 1-105, 276-277; *see also* Intelligence tests
 - abandonment of, by schools, 76, 80
 - bias in, 76, 80
 - in college admissions, 77-78
 - misuses of, 76
 - new directions in research on, 85
 - public controversies, 79-85
 - relevance of, 89-90
 - social effects of, 80-84
 - tasks, 45-46
 - theory, 4, 19, 21, 43-44
 - use for guidance purposes, 76
 - validity, 89-90
- Mental age**, 8, 15-16
 - Binet's scale of, 14
 - concept introduced by Binet, 8
 - variation in, 16
- Mental deficiency**, Binet's studies of, 9
- Mental hygiene movement**, 341, 343, 358-371
 - differences in attitudes of teachers and mental hygienists, 362-365
 - history of, 358-359
 - school courses, 359-362
 - teachers' knowledge of principles, 363-365
 - textbooks, 365-371
- Mental measurement**
 - responsibility of measurement specialists, 84-85
 - responsibility of social scientists, 84
- Mental Measurements Yearbooks (Buros)**, 27, 38, 40, 64, 66, 68-69, 153
- Mental test**, Cattell's introduction of the term, 6
- Mental testing movement**, 3, 457
- Mentalism**, 257-258
- Meritocracy of the intellect**, 79, 81
- Metric scale of intelligence**, 15
- Minimum Foundation School Program (Texas)**, 598
- Misuses of behavioral technology**, 232
- MLAT**, 436
- Modeling**, 237, 245-246
 - approach in language instruction for disadvantaged children, 246

- Models of factor analysis, 55
- Models of interaction between research and language teaching, 407-409
- application model, 408-409, 415, 422, 444
 - convergent model, 409, 433
 - resource model, 409, 422
 - separation model, 408, 415
- Modern Foreign Language Study, 403
- Modern Language Aptitude Test (MLAT), 435-436
- Modern Language Association (MLA), 413, 454-455
- Modern Language Centre, 455
- Moral development
- psychoanalytic view of, 334-336
 - role of superego, 335
- Moray House intelligence tests (Thomson), 10
- Motivating environment, 234
- Multifactor batteries, 40-42
- Multiple correlation, 61
- Multiple-factor analysis, 38
- method, 40
 - theory, 41
- Multiple regression, 61
- Multivariate descriptive statistics, 19-20
- National Board of Education (Sweden), 530
- National Committee on Mathematical Requirements, 177
- National Committee for Mental Hygiene, 358-359
- National Conference of State Legislatures, 617
- National Council of Teachers of Mathematics, 178
- National Defense Education Act (NDEA), 452
- National Education Association (NEA), 148, 477, 599-600
- National Education Finance Project, 617
- National Federation of Modern Language Teachers Associations, 456
- National Foundation for Educational Research (England and Wales), 449, 531
- National Froebel Foundation, 288
- National Institute of Education (NIE), 478-479
- National Intelligence Test*, 8, 154
- National Lawyers' Committee for Civil Rights Under Law, 581, 612, 617
- National Merit Scholarship Qualifying Tests, 88
- National Society for Performance and Instruction, 223
- National Society for Programmed Instruction (NSPI), 223
- National Society for the Study of Education, 148
- National Urban Coalition, 617
- Negative reinforcers, 207
- Negative stimulus, 210
- New curricula, 182, 275-276
- algebra, 177
 - mathematics, 275, 300
- New psychology, 177
- New York Psychoanalytic Society, 322
- New York State Regents Literacy Test, 156
- New York State word list for foreign-born adults, 138
- Newton Readers, 164
- Nominal scales, 60
- Nonlinear factor analysis (McDonald), 55
- Nonlinear models, 51
- Nonparametric statistics, 60
- Nonpromotion in school, 117-118
- Normal probability distribution, 19, 60, 140
- Normal schools, 350
- Normative interpretation of test scores, 18
- Norms for tests, 18
- Norwood Report, 74-75
- Norwood typology, 74-75
- Nuffield Foundation, 453
- Objective (quick-scoring), standardized tests, 7, 114-115
- Objective vocabulary data, 109-122
- Oblique factor structure, 51
- Obscurity factor, 179
- Observational learning, 229, 231, 245, 247
- Odd-even reliability coefficients, 26, 64
- Oedipus complex, 325-327, 364-365, 368, 380
- Officer Qualifying Test (OQT), 40
- Open classroom, 488, 495-497
- Operant behavior, 205-206, 209
- analysis, 252
 - in children, 216
 - in psychotics, 216
- Operant conditioning, 206, 208, 229, 231-232, 242-244, 251, 255
- application of principles, 215-217
- Operant laboratories in pharmacology, 216
- Operant psychology, 216, 224, 252-253
- Operative structures, 301, 302
- Optimality of training conditions, 212
- Oral (speech) vocabularies, 122
- Ordinal scales, 60
- Österåker system of ability grouping (Sweden), 543
- Overlapping adult and child vocabularies, 137-138
- Paper-and-pencil tests, 8, 92
- Papert's LOGO, 295
- Papert's turtles, 307
- Paradigm (Kuhn), 280-282
- Parent-education movement, 379-381
- in popular publications, 380-381
- Parent training in behavior modification, 236
- Path analysis, 61
- Pavlov's theory of conditioned reflexes, 330
- Pearson product-moment correlation coefficient, 20
- Pedagogical grammar, 423-425, 454, 460
- Peer modeling, 249
- Pennsylvania Project, 449-450, 453, 462
- People's Library, 180
- Percent correct score, 18
- Percentage equalizing formulas in school finance, 585
- Percentile (Galton), 18-19
- Perception, 491
- Performance standards, 248
- Permissiveness, 374
- psychoanalysis and, 335-336
- Personal standards, 247-249
- Personality theory, 491
- Personalized System of Instruction (PSI), 225-228
- Phatic language, 431
- Phonetics, 421
- effect on language teaching, 402
- Phonics, 180
- Piaget's concept of developmental stages, 270-271
- sensorimotor stage, 270

- transition between stages, 270-271
- Piaget Conference, 276
- Piaget's empirical method, 269-270
- Piaget's genetic epistemology, 279
- Piaget's views on his theory and practical educational matters, 289-292
- Piagetian conservation tasks, 271
- Piagetian constructive analysis, 300
- Piagetian educational techniques, 292-295
- Piagetian epistemology
 - state, 284
 - structure, 284
 - transformation, 284
- Piagetian literature
 - guidebooks to, 273, 277
- Piagetian tasks, 296-297
- Piagetian techniques, 292
- Piagetian test, 296
- Piagetian theory, 268-278
 - and affective development, 268
 - in curriculum projects, 302-306
 - description of, 283-288
 - and early education, 301-306
 - and educational practice, 267-312
 - epistemology of, 284-288
 - and experimental psychology, 273-275
 - hard core of (Lakatos), 283-288
 - and history of science, 280
 - and Hullian theory, 273
 - implications for education, 288-299
 - influence on child-centered primary schools in England, 288
 - influence on curriculum reform movement in U.S., 288
 - influence of Flavell on, 277-278
 - and intelligence, 298
 - in light of Kuhn's paradigm analysis, 280-282
 - and LOGO, 307-311
 - mathematical applications, 299-301
 - measurement of developmental level, 295-297
 - and new curricula, 276-277
 - and North American psychology, 272-273
 - and philosophy of science, 280
 - postponement of tasks, 304-305
 - rediscovery of, 271-278
 - role of epistemology in, 278
 - and scientific method, 279-283
 - second phase, 269-271
 - specific applications, 299-311
 - stability of developmental level, 295, 297-299
 - third phase, 278
 - and training of mathematics teachers, 301
 - in view of Lakatos' research program, 282-283
 - and writing computer programs, 309-310
- Pigeon training (Skinner), 208-210, 215-217, 222-223, 483
- Pimsleur's Language Aptitude Battery (PLAB), 435-436
- Pinard and Larendau's scale for school-age children, 296
- Planned environments, 238
- Plato's Republic of Kings, 576
- Pleasure principle, 328, 329
- Point scale, 7, 16, 48
- Poll tax, 596
- Positive reinforcement, 207, 210, 237
- Pragmatism, 200
- Pragnanz, 491
- Predictive validity, 64
- Prelateny period in child development, 325-326
- Preoperational stage, 270
- Prescriptive science of educational design, 257
- Presumptive constitutionality, 592
- Primacy of speech, 410, 412
- Primary Education Program, 229
- Primary mental abilities, 42, 52, 89
- Primary Mental Abilities (PMA) battery (Thurstone), 24, 40, 42, 52, 56, 58, 89, 93
- Programmed instruction, 215-224, 226, 228, 255, 434
 - design and implementation, 220-222
- Programmed textbooks, 223
- Progressive education, 113, 173, 182, 289, 341
 - comparison with principles of psychoanalysis, 372-373
 - movement, 5, 120, 320, 371-374, 490
- Progressive Education Association, 371
- Project PLAN, 228
- Project Talent, 41
- Projective tests, 345
- Psychic energy, 383
 - Freud's concept of, 329-330
- Psychoanalysis
 - ambivalence toward, 387
 - in American medicine, 321-322
 - and Catholics, 378
 - in child-development textbooks, 343-350
 - in child-study textbooks, 343-350
 - comparison with principles of progressive education, 372-373
 - cultural impact of, 374-381
 - definition of, 319
 - in educational journals, 341
 - educational movements related to, 340-341
 - elementary education and, 319-388
 - in general psychology textbooks, 323
 - in general publications, 341-342
 - Hall's interest in, 342
 - history of, 320-324
 - impact on education, 340-381, 386-387
 - Lashley's views on, 323
 - mental hygiene and, 386
 - nonprofessional channels of influence, 387
 - parts of theory relevant to education, 324-340
 - in pedagogical textbooks, 341
 - and permissiveness, 334-336
 - in psychology, 322-324
 - sexual emphasis, 354-355
 - theory of learning, 327-331
 - Thurstone's views, 322-323
 - view of limitations of education, 388
 - view of teacher's role, 336-340
- Psychoanalytic theory, usefulness to teachers, 337-340
- Psycholinguistics, 113, 405-407, 418, 459
 - impact on second-language education, 460-461
- Psychological measurement theory, 4, 14, 22-23
 - axiomatic formulation, 60-61
- Psychological measurements, 2, 8, 22-23, 60
- Psychological utilitarianism (Thorndike), 385
- Psychology
 - association with education, 201-204
 - as an experimental science, 202
 - laboratory, 199
 - relationship to education, 199-204
 - scientific, 199

- Psychometric Society, 3-4, 37
 Psychometric test of mental ability, 374
 Psychometrics, 1, 37, 44-46, 403, 413
 methods, 15, 435
Psychometrika, 3-4
 Psychophysics, 5, 15, 23
 Psychosexual stages, 327
 Psychosocial research, 436
 Punishment, 207
 as means of education, 328-329
 Purposeful reading, 121
- Q** (Quantitative) scores, 32
 Quarks (Feynman), 507
 Quiz kids, 481
- Rat training** (Skinner), 212
 Ratio scales, 60
 Ratio schedule, 206
 Raw score, 18
 Reaction formation, 326
 Readability, 130, 132, 163-164
 of arithmetic materials, 155
 conceptual network of, 179-182
 cumulative properties, 178-182
 of governmental publications, 111
 measures, 125, 153
 personal-historical network of, 79-182
 research, 108-109
 of test instructions, 155
 Readability formulas, 111, 126, 139, 153, 160, 172, 179
 application to textbooks, 159-161
 use by publishers of psychological tests, 153
 use by textbook publishers, 153
 Readability Laboratory at Teachers College, 180
 Reading
 comprehension, 119, 121, 179
 difficulty, 332
 instruction, 120
 materials and teacher competence, 158
 for meaning, 120
 vocabulary, 115, 121-122
 Reading disability, 118, 332
 emotional factors, 332-333
 Reading matter, word range of, 131
 Reading Vocabulary for the Primary Grades, 136
Realexamen, 552
 Reality principle, 328-329
Realskola, 532, 535-536, 539-540, 542-547, 551-552, 558, 561, 564-565, 569-570
 Recognition (hearing) vocabulary, 113, 121
 Rectangular classroom, 488-490
 Regional English Language Centre, 456
 Regression (statistical), 20
 Reinforcement, 206, 213
 and cognition, 251-252
 intermittent, 206-207
 role in learning, 328
 schedules, 206-207
 theory, 213, 234
 Reinforcer, 206-208, 234
 Rejall reading vocabularies, 138
 Reliability coefficients, 21-22, 26, 38, 46, 48
 Reliability of tests, 38, 46, 48
 equivalent form, 26
 odd-even, 26, 64
 split-half, 29
 theory of, 21-22
 Report on Vocabulary Selection, 139
 Representational systems, 246-247
 Repression, 329-332, 337
 Research; *see also* Educational research and Basic research
 definition of, in second-language education, 399-401
 history of, in language teaching, 401-407
 impact on preparation of educational administrators, 498
 linguistics and sociolinguistic influences in second-language education, 410-433
 models of interaction with language teaching, 407-409
 psychological and psycholinguistic influences in second-language education, 433-448
 Research program (Lakatos)
 hard core and protective covering, 282-288
 as synonym for theory, 282-288
 Reserves of ability, 525, 527, 535; *see also* Talent reserve
 assessment of, 531-541
 Reserves of talent, 527, 535; *see also* Talent reserve
 Resource model, 459
 Respondent conditioning, 205, 231
 Response maintenance, 232-234
 Rewards as means of education, 328-329
Robinson v. Cahill, 616-617
Rodriguez case, 594
 and school-finance reform, 581-620
 summary of, 582-583
Rodriguez v. San Antonio, 582, 591-606
 Rotation of factors to simple structure, 51, 53-55
 Royal Commissions (England)
 Plowden Committee, 575
 Robbins Committee, 575
 Royal Commissions (Sweden), 523-531
 Russell Sage Foundation, 80
 Russian word list, 124
- S-R**
 bonds, 201
 reflex, 205
 reinforcement theory, 213
 Sampling theory of intelligence, 10, 62
 San Antonio, school finance in, 582
San Antonio v. Rodriguez, 581
 interaction of law and social science, 591-606
 Satiation, 237
 Scaling of tests, 17
 theory, 3, 14-18, 59
 Thurstone's absolute scaling methods, 23
 Schedules of reinforcement, 216, 218
 Scherer-Wertheimer language-teaching experiment, 448-449, 453
 Scholastic aptitude, 9, 67
 Scholastic Aptitude Examinations, 9, 42-44; *see also* Scholastic Aptitude Test (SAT)
 Scholastic Aptitude Test (SAT), 38, 42-44, 67, 77-78
 decline in scores, 82
 Scholastic aptitude tests, 36, 38, 62, 77, 80, 116
 effects of coaching, 78
 School administrators, 504; *see also* Educational administrators
 School differentiation in Sweden, 529

- School failure, 117-119
 - role of reading in, 118-122
- School finance
 - in California, 583, 590, 616-617
 - in Connecticut, 604, 616-617
 - developments in public policy on, 615-620
 - developments in social advocacy on, 615-620
 - fiscal bias in Texas system, 595
 - foundation plan (Strayer and Haig), 585
 - full state assumption, 585-586
 - in Hawaii, 582, 604
 - in Illinois, 589
 - influence of economics, 586-588
 - in New Jersey, 616-617
 - percentage equalizing formulas, 585
 - policy research on, 615-620
 - recent legal developments, 615-617
 - in San Antonio, Texas, 582
 - state aid and, 585-588
- School-finance litigation
 - application of equal-protection analysis to, 593
 - attorneys and social scientists as partners in advocacy, 612-615
 - compelling state interest, 603-604, 608-611
 - definition of equal education, 602-603
 - judicial reasoning, 606-612
 - role of social science in, 600, 606-612
 - sources of evidence, 606-607
- School-finance reform
 - contemporary legislation, 617
 - cost-quality question, 600-601
 - criterion of relevance, 607-608
 - District Power Equalizing, 590-591
 - educational issues in, 593-606
 - First Amendment rights, 600
 - judicial resolution of educational issues in, 593-606
 - the law and, 581-620
 - legal scholarship, 588-591
 - Serrano* case, 588
 - use of federal constitutional litigation in, 581-583, 591-620
 - use of federal courts in, 591
- School-finance research, 583-591, 617-619
 - allocation of government and private resources, 587
 - Cubberley's contributions to, 583-585
 - eclecticism in, 584-591
 - effects of equal-protection litigation on, 584
 - fiscal capacity, 587
 - orientation toward advocacy, 583-591
 - public expenditure models, 587-588
 - taxation, 587
 - use of computer-simulation techniques, 584
 - use of empirical analysis, 584
 - use of production-function analysis, 588
- School finances, fiscal discrimination in, 596
- School performances, changes in, 116-117
- Schultz's studies on economics of education, 487
- Science Curriculum Study (Karplus), 485-486
- Scientific management, 500-501, 503
 - applications to educational administration, 498-499
- Scientific theory
 - Dewey's views of its relation to practical applications, 201-202
 - Thorndike's views of its relation to practical applications, 201-202
- Scoring keys, 24
- Screening in school admissions (Sweden), 550
- Second-language acquisition
 - attitudes and motivation as factors in, 436-438
 - built-in syllabus of learner, 442, 445
 - cognitive strategies, 446
 - critical age for, 439
 - description of successful classroom learning, 447
 - empirical studies, 444-445
 - environmental factors, 447
 - error analysis, 443, 445-446
 - immersion program, 441
 - incentive programs, 437-438
 - individual differences in language-learning processes, 446-447
 - instrumental orientation as factor in, 436-437
 - integrative orientation as factor in, 436-437
 - interlanguage, 445-446
 - language acquisition device (LAD), 442
 - language aptitude as a factor in, 434-436
 - mental processes involved in, 441-447
 - optimal age for, 438-441
 - psychological variables, 434-441
 - psychosocial factors, 436-438
 - relationship to first-language acquisition, 445
- Second-language education
 - behavioristic model, 433-434
 - Chastain study, 449
 - explicit and implicit language teaching, 450
 - guidelines derived from psychological theory, 444
 - GUME Project, 449-450
 - impact of descriptive linguistics, 460
 - impact of educational research, 462
 - impact of language sciences, 397-465
 - impact of linguistic theory, 458-460
 - impact of psychology and psycholinguistics, 460-461
- Lambert immersion studies (Canada), 449-450
- linguistic and sociolinguistic influences, 410-433
- mechanisms for information exchange with research, 453-457
- mechanisms for initiating research and development, 451-453
- mechanisms for interaction with research, 451-457
- National Foundation for Educational Research study (Britain), 449
- NDEA Language Development Program, 452-453
- Pennsylvania Project, 449
- psychological and psycholinguistic influences, 433-448
- teacher training, 457
- Thorndike studies, 448-449
- Second-language learning
 - psychology of, 403-404, 406-407
 - relation between age and language teaching, 574
- Sectioning, 543
- Secure test, 67
- Selective schools, 75-76, 527-528
 - in Britain, 79
- Self-concept studies (Clark), 487
- Self-control, 215, 232, 242-245, 252; *see also* Self-management
- Self-determination, 241-244
- Self-esteem and resulting achievement, 248

- Self-evaluative processes, 248
 Self-instructional test-scoring devices, 222
 Self-management, 244, 251; *see also* Self-control.
 of instruction, 228
 Self-modification, 250
 Self-monitoring, 248-250
 Self-presented consequences, 243
 Self-regulation, 241-245, 248, 250-251; *see also*
 Self-control.
 Self-reinforcement, 247-252
 Semantic Count, 126, 151, 154
 Semantic problem, 151
 Semantic structure, 425
 Semantics, 407, 424
 componential analysis, 426-427
 functional analysis, 428
 structural analysis, 425-426
 and the teaching of grammar, 427-428
 and vocabulary teaching, 425-427
 Sensory deprivation (Hebb), 487
 Sequential Tests of Educational Progress (STEP),
 155
 Serrano case, 588, 594, 612-617
 Serrano v. Priest, 583, 590
 Service words, 150
 Sexual instincts, 383
 Shaping, 217, 245
 in language learning, 442
 of responses, 208-209, 216
 Sight words, 149
 Sign-word, 147
 Silent reading, 120
 Simplified primer, 124
 Skinner's pigeon experiments, 483
 Skinner's system
 application to classroom, 217-220
 of behavior, 204-206
 and the classroom, 218
 counter influences, 212-214
 Skinner as Visiting Father, 217
 Small sample statistical methods, 20
 Social advocacy, recent developments, 615-620
 Social composition of academic secondary schools
 in France, 551
 in Soviet Union, 551
 in Sweden, 551-555
 Social Darwinism, 5
 Social factors in school differentiation (Sweden),
 549-555
 Social handicaps
 effects on school careers of Swedish children,
 527-528
 in selection by schools (Sweden), 551-555
 Social learning theory, 244-248, 251-252
 Social reinforcement, 238
 Social science and the law in school-finance
 reform, 581-620
Social Sciences Citation Index, 126
 Sociolinguistic analysis, 430
 Sociolinguistics, 407, 410-433, 459
 effect on language teaching, 430-433
 Sociometric structure, 492
 Sociometry, 501
 Spearman-Brown prophecy formula, 21, 26
 Spearman-Holzman Unitary Trait Study, 23
 Spearman's two-factor theory of intelligence,
 12-13, 23
 Specialized word knowledge, 115
 Specific factor in intelligence, 12, 51
 Speech community, 432
 Speech event, analysis of, 430-433
 Speed perception, 444
 Speededness of tests, 26, 29, 56, 64, 76, 86
 Spelling
 conscience, 128
 curriculum, 145
 lists, 128, 159
 vocabularies, 125
 Split-half reliabilities, 29
 Spoken Language Training Program of the
 Foreign Service Institute, 405
 Sputnik, 320
 effect on curriculum reform, 275
 Square classroom, 488, 491-492
 Stage theory, 295-299, 301, 306
 Kohlberg's, 298-299
 Standard scores, 18
 Standardization of tests, 4, 8, 24-33, 42, 62
 Standardized achievement tests, 115-116
 effects of vocabulary studies on, 152-156
 Standardized reading materials, 124
 Stanford-Binet scale, 7, 22, 29
 Stanford-Binet test (Terman), 15, 47
 State purpose in school-finance litigation, 592-593
 Statistical methods, 22, 61-62
 and mental test theory, 4, 18-24, 38
 Statistical weights, 61
 Statistics, 22, 61-66
 bivariate descriptive, 19-20
 inferential, 19-20
 multivariate descriptive, 19-20
 univariate descriptive, 18-19
 Sternberg's componential analysis of mental tasks,
 59
 Stevens' scale type, 60
 Stimulation-deprivation studies, 494-495
 Stimulus abstraction, 210-212
 Stimulus control, 216
 Stimulus generalization, 210-212
 Stimulus-response bonds, 201, 489
 Stimulus-response theory, 201, 204-205, 330
 Strict judicial scrutiny, 599
 in school-finance litigation, 591-592
 Structural analysis, 426
 Structural linguistics, 407, 410-414, 421
 Bloomfield's, 404-405
 Structural model of the mind (Freud), 324, 336
 Structural semantics, 425-426
 Structuralism, 415, 424
 Structure of Intellect (SI) Model (Guilford), 41,
 57-58, 85, 486-487
 "Student's" small-sample statistical methods, 20
 Student Welfare Commission (Sweden), 532
Studentexamen, 536, 538
 Studies in word usages, 122-124
 Study committee of the International
 Kindergarten Union, 127
 Stuttering, 126
 Sublimation, 326, 328-329, 332
 Substitution, 237
 Successive approximations, 208-209, 219, 237,
 245
 Summerhill, 335-336
 Superego, 335

- Supreme Court and school-finance reform, 582-583, 591-606
- Suspect class, 596-597
in school-finance litigation, 591-592
- Swedish Employers Association, 528
- Symbolic processes, 247
- Systems analysis, 224
- T-scores, 18
- Tailored tests, 47, 62
- Talent reserve, 549
actual, 538
broad and narrow interpretations of, 538
estimation of, 535-541
Härnqvist study, 539-541
in lower social strata, 533-535
potential, 538, 540
probability method (Härnqvist) used to study, 539-541
- Task analysis, 256
- Task homogeneity, 17
- Teacher
and children's disguised sexual curiosity, 338
contrast to therapist, 337-340
knowledge of mental hygiene principle, 363-365
personal psychoanalysis of, 337
psychoanalytic view of, 336-340
- Teacher's manuals, 163
- The Teacher's Word Book*, 108, 110, 113-115, 123, 125-127, 403
compilation of, 126
use in foreign word books, 148-149
- Teachers College, vocabulary research tradition at, 180-182
- Teaching
styles, 492
of vocabulary, 119
- Teaching English to Speakers of Other Languages (TESOL), 414
- Teaching machines, 204, 215-224
design and implementation, 220-222
- Technical vocabularies, 109, 129-134, 165
- Technical word, 150
- Terman Group Test of Mental Ability*, 8
- TESOL Association, 455
- Test
administration, 62
cloze, 6, 179
completion, 6
computer-administered, 47
construction, 4, 24-33, 44-46, 62, 153, 155
experimental sections, 42
item analysis of, 22
memory-span, 11
methodology, 24
norm, 18
reliability of, *see* Reliability of tests.
scaling of, 17-18
security, 43, 67
speeded, 26
standard score for, 18
standardization of, 4, 8, 24-33, 62
T-score for, 18
use of, 33-37
validation of, 4, 8, 24-33, 62
- Test anxiety, covert self-control and, 250
- Test of fiscal neutrality theory (school-finance reform), 590
- Test item
cycle-omnibus form, 63
difficulty, 47
weighting of, 47
- Test score, 18, 36
biased, 36
normative interpretation of, 18
- Test theory, 15, 22-23, 46-50, 86, 89, 92-93
classical, 46
- Test vocabulary, 153
Thorndike's impact on, 153-155
- Testing movement, 114
experimental laboratory phase, 16
- Textbook, 129-134
censorship, 159
evaluation, 143
size, 161-162
supplements to, 162-163
- Theories of intellectual abilities, 44-46
- Theory of testing, 385
- Thinking games, 302
- Thorndike's laws of learning, and psychoanalysis, 330
- Thorndike's views on elementary education, 383-385
- Thorndike vocabulary test, 129
- Thorndike-Woodworth studies on doctrine of formal discipline, 483-485
- Three-mode factor model (Tucker), 55
- Thurstone's PMA studies, 53, 58, 86
- Thurstone's *Psychological Examination for College Freshmen*, 8
- Token economy, 233, 239
in an institutional setting, 238
in mental hospital, 233-235
- Token reinforcement, 240-241
in the classroom, 235
evaluation of, 240-241
programs, 238, 239
- Topographic model of the mind (Freud), 324
- Tracking, 611
- Trait theory, 230-231
- Transfer of training, 232, 483
Thorndike-Woodworth studies, 483-485
- Transformational generative grammar, 407, 414-421, 424, 442
Chomsky's model of, 414-415
competence, 416-417
deep structure, 416, 418-419
L2 materials and, 418-420
L2 teaching methodology and, 420
performance, 416-417
surface structure, 416, 419-420
- Transformational generative linguistics, 430
- Tripartite school system, 74
- TTWB index number, 132
- Twins studies (Burt), 88
- Two-factor theory of intelligence (Spearman), 12, 23
- Type I errors in hypothesis testing, 20
- U.S. Office of Education, 617
- UCLA Psycholinguistics Conference, 433-434, 453
- UNESCO Institute for Education, 440
- Unig (Walker), 17
- Unitary Trait Study, 13
- Univariate descriptive statistics, 18-19

- Validation of tests, 4, 8, 24-33, 42, 62
- Validity of test, 26, 38, 89-90
 - concurrent, 64
 - predictive, 64
- Vander Beke frequency list, 149
- Varimax method for orthogonal factors, 55
- Verbal encoding, 246
- Verbal models, 247
- Verbalism, 113
- Vicarious learning, 245
- Vicarious reinforcement, 229, 246, 248, 251
- Vocabularists, 108, 138, 144, 152, 160
- Vocabulary
 - annual growth, 109
 - control, 160
 - cultural variability, 147-148
 - hearing (recognition) vocabulary, 134-137
 - overlap of children's and adults', 134-139
 - semantics and the teaching of, 425-427
 - speech (oral), 134-137
 - writing, 134-137
- Vocabulary building, 145-146, 152
 - program, 132
- Vocabulary burden, 132, 144
- Vocabulary counters, 110
- Vocabulary development of the American child, 111-114
- Vocabulary distribution, principles of, 164
- Vocabulary knowledge, 153
 - precision, 154-155
 - range, 154-155
- Vocabulary load, 160, 163
- Vocabulary research, 107-198
 - algebra and, 177-178
 - conceptual network of, 179-182
 - cumulative properties, 178-182
 - effect on educational practice, 165-183
 - effect on schools, 152-183
 - in foreign-language teaching, 168-169
 - geography and, 175-177
 - in literature teaching, 170-172
 - personal-historical network of, 179-182
 - in reading, 167-168
 - in shorthand teaching, 169-170
 - in science education, 130-131
 - social sanction for, 166-167
 - sociological factors and, 172-178
 - in teaching the fine arts, 170
 - in teaching of typing, 170
 - textbook editors' use of, 164
 - of textbooks, 130-134
 - in word fields, 167-172
- Vocabulary size, 100-101, 113
- Vocabulary teaching, 165
- Vocabulary test-makers, 114
- Vocabulary tests, 115, 119
- Weighted Median Index Number, 132
- Weighted scores, 47, 51
- Winnetka Plan, 228
- Woods Hole Conference, 276, 277
- Word analysis, 121
- Word counters, 114, 122-124
- Word counts, 108, 115, 125, 179
 - of scientific terms, 131
 - use of dictionary for, 110
- Word difficulties, 163
- Word frequency, 115
 - cross-national equivalents, 148-149
 - foreign language, 148-149
 - geographical variability, 147
 - implications for education, 134-152
 - as measure of importance, 144
 - reliability of word placement, 142
 - skewed distribution curve, 140-146
- Word-frequency list, 126
- Word-frequency studies, 108
 - before 1921, 122-125
 - of newspapers, 124
 - use in literacy tests, 156
 - use for shorthand, 124
 - use in textbook publication, 156-165
- Word handbooks, 122
- Word knowledge, 110, 114, 119, 129; *see also* Vocabulary size.
- Word lists, computers and, 127
- Word meaning, 150-151
- Word power, role of, in reading, 119
- Word range, 146
 - as measure of importance, 144
- Word recognition, 121, 151
- Word selection, 142
- Words, 108, 124
 - analysis of meaning, 42
 - characteristics of frequent, 146-152
 - culture-fair, 147
 - difficulty of, 129-130
 - durability of, 149-150
 - form(ative), 146
 - function of, 146
 - grammatical, 146-147
 - importance of, 129-130
 - knowledge of, 113
 - lexical, 147
 - mastery of, 140
 - meaning load, 152
 - multiple-meaning, 150-152
 - scientific connotation of common, 130
 - service, 146
 - technicality of, 129-130
 - use of, 113
- Workbooks, 163
- Writing vocabularies, 115
- X-rays (Roentgen), 386-387
- Yearbooks of the National Society for the Study of Education, 156
- Zero-value words, 132